

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 20.31 seconds
(without alignments)
1095.470 Million cell updates/sec

Title: US-09-545-072A-2
Perfect score: 1922
Sequence: 1 MEAYKQWNRREYVQSGS.....EKIVLEGAQSRITYISGS 367

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	100.0	367	21	Arabidopsis thalia
2	249.5	13.0	336	21	Human vesicle asso
3	186.5	9.7	391	21	Yarrowia lipolytic
4	89.5	4.7	956	22	Corynebacterium gl
5	89.5	4.7	1257	17	Alpha-ketoglutaric
6	89.5	4.7	1257	19	B. lactofermentum
7	89	4.6	3165	14	Sequence encoded b
8	88	4.6	1645	18	B. pertussis adeny
9	88	4.6	1705	10	Sequence of part o
10	88	4.6	1706	11	Adenyl cyclase fro
11	87.5	4.6	302	20	HIV Tat protein.

12	87.5	4.6	302	22	AA887879	Synthetic fusion p
13	87	4.5	1522	10	AA933357	Sequence of the ca
14	86.5	4.5	197	21	AA952858	Arabidopsis thalia
15	86.5	4.5	288	21	AA952858	Arabidopsis thalia
16	86.5	4.5	298	21	AA952857	Arabidopsis thalia
17	86.5	4.5	1025	17	AA903185	Rice Xa21 disease
18	86.5	4.5	1025	20	AA903185	O. longistaminata
19	86.5	4.5	1025	22	AA967442	Amino acid sequenc
20	86.5	4.5	1026	20	AA939279	Xa21 signal transd
21	85.5	4.4	2380	21	AA181315	Plasmodium falci
22	85.5	4.4	2595	20	AA939297	spnA a polyketide
23	85	4.4	424	20	AA935407	Chlamydia pneumoni
24	84.5	4.4	1021	19	AA976192	Actinoplanes sp. a
25	84.5	4.4	4150	21	AA92707	S. antibioticus 8,
26	84	4.4	580	21	AA982157	Pseudomonas putida
27	84	4.4	765	20	AA989561	Recombinant heat-r
28	83.5	4.3	309	21	AA938792	Arabidopsis thalia
29	83.5	4.3	460	21	AA938791	Arabidopsis thalia
30	82.5	4.3	177	21	AA924979	Arabidopsis thalia
31	82.5	4.3	197	21	AA924978	Arabidopsis thalia
32	82.5	4.3	288	21	AA924977	Arabidopsis thalia
33	82	4.3	220	21	AA916938	Arabidopsis thalia
34	82	4.3	248	21	AA916937	Arabidopsis thalia
35	82	4.3	252	21	AA916936	Arabidopsis thalia
36	80.5	4.2	309	21	AA917027	Arabidopsis thalia
37	80.5	4.2	460	21	AA917026	Arabidopsis thalia
38	80.5	4.2	524	15	AA956525	Protein kinase (NU
39	80.5	4.2	524	16	AA976621	Saccharomyces cere
40	80.5	4.2	747	20	AA989585	Human ATP-binding
41	80.5	4.2	747	21	AA913355	Human ATP-binding
42	80.5	4.2	752	20	AA914068	Human ABC-Transpor
43	80	4.2	400	21	AA959923	Arabidopsis thalia
44	80	4.2	401	21	AA959922	Arabidopsis thalia
45	80	4.2	730	16	AA982881	Lupin exo-(1-4)bet

ALIGNMENTS

RESULT 1
AAB19718
ID AAB19718 standard; Protein; 367 AA.
XX
AC AAB19718;
XX
DT 19-FEB-2001 (first entry)
XX
DE Arabidopsis thaliana SSE1 protein.
XX
KW SSE1; shrunken seed gene; storage reserve; storage protein;
oil body; transgenic plant.
XX
OS Arabidopsis thaliana.
XX
FH Key
FT Region Location/Qualifiers
FT Region 38..53
FT Region /note= "hydrophobic region"
FT Region 109..124
FT Region /note= "hydrophobic region"
FT Region 134..180
FT Region /note= "hydrophilic region"
FT Region 201..227
FT Region /note= "hydrophilic region"
FT Region 238..257
FT Region /note= "hydrophobic region"
FT Region 337..355
FT Region /note= "hydrophobic region"
XX
PN WO200061735-A1.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09192.

```
XX PR 08-APR-1999; 99US-0128651.
XX (GEO ) GEN HOSPITAL CORP.
XX Lin Y;
XX WPI: 2000-679483/66.
XX N-PSDB; AAA88782.
XX Novel shrunken seed gene useful for producing transgenic plants having
XX altered production of food storage reserve material, intracellular
XX transport of storage protein and formation of protein or oil bodies
XX Claim 2; Page 58; 64pp; English.
XX The present sequence is that of Arabidopsis thaliana SSE1 (shrunken
XX seed) protein, as deduced from isolated cDNA (see AAA88782). SSE1
XX protein, when expressed in a cell of a plant, modifies or alters
XX the production of a food storage reserve material (e.g. protein,
XX lipid or carbohydrate storage reserve), facilitates the
XX intracellular transport of a storage protein, or facilitates the
XX formation of protein or oil bodies. The invention provides a
XX transgenic plant (or plant cell, plant tissue, plant organ or
XX plant component) which includes a recombinant SSE1 transgene that
XX modifies the production of food storage reserves, thereby
XX increasing nutritional value. An antisense construct is useful for
XX modifying desiccation tolerance.
XX Sequence 367 AA;
XX
Query Match 100.0%; Score 1922; DB 21; Length 367;
Best Local Similarity 100.0%; Pred. No. 6.2e-195; Mismatches 0; Gaps 0;
Matches 367; Conservative 0; Indels 0;
Qy 1 MEAYKQWVRNREYVQSGFANGLTWLPKFSASEIGPEAVTAFIGFTTINEHIEN 60
Db 1 meaykqwwrrnreyvqsgsfangltwlpkfsaseigpeavtafigfttinehi 60
Qy 61 APTPRGHVGGSGNDPSLSYPLLIILAILKDLTVVVEVAEHFYGDKKWNYIILTEAMKAVIR 120
Db 61 aptprghvsgsgndpslsyplliailkdlctvvevaeahfygdkkwnyilteamkavir 120
Qy 121 LALFRNSGYKMLLOGGETPNEEKDSNQESQNRAGNSGRNLGPHGLGNQNHHPNLEGR 180
Db 121 lalfrnsykmllggetpneekdsngsqesqnragnsgrnlghpqlgnqnhhpwnlegr 180
Qy 181 AMSALSSFGQNARTTSTTPGWSRRIOHQQAVIEPPMIKERRRTMSSELLTEKGVNGALFA 240
Db 181 amsalssfgqnarttsttpgwsrriohqqavieppmikertrmsellselltekgvngalfa 240
Qy 241 IGEVLYITRPLTYLVLFIRKYGVRSWIPWALSISLSDVTGLMGLLANSKWKGEKSKOVHESGP 300
Db 241 igevlyitrpltylvlfirkygvrswipwalsisvdtlmgllanskwgkkskvhsfsgp 300
Qy 301 EKDELRRRLKIWALYLMRDPFTTYTROKLESSQKKLELIPFLIGFLTEKIVELLEGAQSR 360
Db 301 ekdelrrrllkiwalylmrdpfttytroklessqkklelpliflgtelkivellegasqr 360
Qy 361 YTYISGS 367
Db 361 ytyisgs 367
RESULT 2
AAB24227
ID AAB24227 standard; protein; 336 AA.
XX AC AAB24227;
XX AC AAB24227;
XX DT 07-FEB-2001 (first entry)
XX
```

```
DE Human vesicle associated protein 6 SEQ ID NO:6.
XX Human; vesicle associated protein; VEAS; diagnosis; nootropic;
XX immunosuppressive; antibacterial; antihelminthic; protozoacide;
XX virucide; anti-inflammatory; cytostatic; gene therapy; akinesia; AIDS;
XX transport disorder; Bell's palsy; diabetes mellitus; catatonia;
XX infectious myositis; dementia; Grave's disease; autoimmune disorder;
XX inflammatory disorder; asthma; bronchitis; Reiter's syndrome; cancer;
XX systemic lupus erythematosus; rheumatoid arthritis; fungicide;
XX infection.
XX Homo sapiens.
XX OS WO200060082-A2.
XX PN 12-OCT-2000.
XX PD 06-APR-2000; 2000WO-US09353.
XX PF 07-APR-1999; 99US-0128193.
XX PR 20-JUL-1999; 99US-0144701.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Lal P, Yue H, Hillman JL, Baughn MR, Tang YT, Lu DAM, Azimzai Y;
XX WPI: 2000-665006/64.
XX DR N-PSDB; AAC60990.
XX PT New vesicle associated proteins (VEAS) and polynucleotides encoding
XX VEAS for diagnosing, treating or preventing transport disorders (e.g.
XX Bell's palsy), autoimmune/inflammatory disorders (e.g. AIDS) or cancer
XX (e.g. leukemia)
XX Claim 1; Page 84-85; 112pp; English.
XX
AAC60995 to AAC61000 and AAC64367 to AAC64369 encode the human vesicle
associated proteins 1 to 19 (VEAS-1 to 19) given in AAB24222 to AAB24240.
XX Human VEAS proteins have nootropic, immunosuppressive, antibacterial,
XX antihelminthic, protozoacide, virucide, anti-inflammatory and
XX cytostatic activities. The isolated polypeptides and polynucleotides are
XX useful for diagnosing, treating or preventing transport disorders (e.g.
XX akinesia, Bell's palsy, diabetes mellitus, catatonia, infectious
XX myositis, dementia, Grave's disease or AIDS), autoimmune/inflammatory
XX disorders (e.g. AIDS, asthma, bronchitis, Reiter's syndrome, systemic
XX lupus erythematosus, rheumatoid arthritis, or viral, fungal, bacterial,
XX protozoal or helminthic infections), or cancer (e.g. adenocarcinoma,
XX leukemia, melanoma, sarcoma, or gall bladder, heart, kidney, bone,
XX brain, breast, pancreas or skin cancers). The polypeptides may also be
XX used to immunise a host animal. The polypeptides and polynucleotides are
XX further useful for screening libraries of compounds or for drug
XX screening assays.
XX Sequence 336 AA;
XX
Query Match 13.0%; Score 249.5; DB 21; Length 336;
Best Local Similarity 25.4%; Pred. No. 8.2e-18; Indels 51; Gaps 12;
Matches 94; Conservative 68; Mismatches 157;
Qy 4 YKQWVRNREYVQSGFANGLTWLPKFSASEIGPEAVTAFIGFTTINEHIENAPT 63
Db 11 yqevytrhpaaataqlatavrgfsyllagrfadshelselvysasnllvldngilrkelr 70
Qy 64 PRGHVGGSGNDPSLSYPL---IAILKDLTVVVEVAEHFYGD-KKWNVIILTEAMKAVI 119
Db 71 kklpv-----slsqgklltwlsvlecvevfmemgaakvwgvgvrlvalqlakavl 123
Qy 120 RLALFRNSGYKMLLOGGETPNEEKDSNQESQNRAGNSGRNLGPHGLGNQNHHPNLEGR 179
Db 124 rmlll--lwfakglq--tppivpldretaqppdgd-----hsqpnhe- 163
Qy 180 RAMSALSSFGQNARTTSTTPGWSRRIOHQQAVIEPPMIKERRRTM--SELLTEKGVNGA 237
```

```

Db 164 --qsyvkrsvrvrtlqntcp--slhsrhwga----pqqregrqgqdhcheelsatptplgl 215
Qy 238 LFAIGEVLVYTRPLIYVLFIRKYGVRSWIPWAIWSLSDVTLGMLLANSKWWGKSKQVHF 297
Db 216 qetiaeflyiarplhlslglsqgrswkpwllagvvdvtslsllsdrkgltr----- 269
Qy 298 SGPEKDELRRKLIWALYMRDFFTKYTRQK-LESSQKLELIPLIGLFLTEKIVELLE 356
Db 270 ---errelrrtlllyllrsfydrfsearillfillqlladhvpvgvltvtrplmdyldpt 326
Qy 357 AQRSTYVISG 366
Db 327 wqkiyfywsg 336

RESULT 3
AAB19719
ID AAB19719 standard; Protein; 391 AA.
XX
AC AAB19719;
XX
DT 19-FEB-2001 (first entry)
XX
DE Yarrowia lipolytica Pexl6 protein.
XX
KW SSE1; shrunken seed gene; storage reserve; storage protein;
KW oil body; transgenic plant.
XX
OS Yarrowia lipolytica.
XX
PN WO200061735-A1.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09192.
XX
PR 08-APR-1999; 99US-0128651.
XX
PA (GEO) GEN HOSPITAL CORP.
XX
PI Lin Y;
XX
PI WPI; 2000-679483/66.
XX
DR Novel shrunken seed gene useful for producing transgenic plants having
PT altered production of food storage reserve material, intracellular
PT transport of storage protein and formation of protein or oil bodies -
XX
PS Disclosure; Page 59-60; 64pp; English.
XX
CC The present sequence is that of Yarrowia lipolytica Pexl6 protein.
CC The invention includes a method for modifying or altering the
CC biosynthesis of a storage reserve in a transgenic plant (or its
CC component) by: introducing a transgene encoding an SSE polypeptide
CC linked to a promoter functional in plant cells; and regenerating a
CC transgenic plant from the transformed cells. The SSE polypeptide
CC is expressed in the cells of the transgenic plant, thereby
CC modifying or altering the seed storage reserve (lipid, storage
CC protein, or carbohydrate e.g. starch) of the plant. In preferred
CC embodiments, the expressed polypeptide is Pexl6.
XX
SQ Sequence 391 AA;

Query Match 9.7%; Score 186.5; DB 21; Length 391;
Best Local Similarity 20.7%; Pred. No. 4.9e-11;
Matches 86; Conservative 68; Mismatches 143; Indels 119; Gaps 13;

Qy 1 MEAYKQWNRNRYVQSGFSGFANGLTWLLPEKFSASEIGPEAVTAFPLGFTTINHHIEN 60
Db 20 ldsydkflvrnaasigsistlrtsyvlpgpfndveiatetlyavlrvlgyhdtliar 79

```

```

Qy 61 APTPRGHVSGNDPSLSYP-----LLIAILKDLTVVVEAAEH 99
Db 80 a-----vaasnaaavypspbnrytdwfiknrkgykyaasravtfvkgelvaemvakk 133
Qy 100 FYGD-KKNYIILTAMKAVIRLALFRNSGYKMLLOGGETPNEEK----- 143
Db 134 nggemarwkciigiegikagiriymlgstlyqplc---ttypdrevtgelleticrdeg 190
Qy 144 ---DSNQSESONRAGSNRNLGPHGLGNQHNHPWNLGRAMSALSLSFGQNARTTTSST 199
Db 191 eldiekgldmpqwknpnrtgrtipeia-----ptnvegylt----- 226
Qy 200 PGWSRRHQHQAVIDEPPMIKERRRTMSELLTEKGNGALFAIGEVLVYTRPLIYVLFIRK 259
Db 227 ----kvlsedvdrpynllsr-----ldnwgy-----vaellsilrplyacilfr 268
Qy 260 YGVR-----WIPWAIISLSDVTLGMLLANSKWWGKSKQVHSGP--- 300
Db 269 qhvnkvtvpastkskfplnspwapwllglviealsrkmmgs---wllrrrq---sgktp 322
Qy 301 --EKDELRRLKIWALYLMRDPFFTKYTRQKLESSQKLELIPLIGLFLTEKIVELL 354
Db 323 aldmqevkgrtnllgwlfrgefgyaytrpillysivarlekigpdlgfgalisdyl 378

RESULT 4
AAB79523
ID AAB79523 standard; Protein; 956 AA.
XX
AC AAB79523;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:562.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100844-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.

```


RESULT	6
AAW41781	
ID	AAW41781 standard; Protein; 1257 AA.
XX	
AC	AAW41781;
XX	
DT	11-JUN-1998 (first entry)
XX	
DE	B. lactofermentum alpha-ketoglutaric acid dehydrogenase.
XX	
KW	Alpha-ketoglutaric acid dehydrogenase; recombinant production;
KW	L-glutamic acid.
XX	
OS	Brevibacterium lactofermentum.
PN	WO9748790-A1.
XX	
PD	24-DEC-1997.
XX	
PF	04-JUN-1997; 97WO-JP01886.
XX	
PR	17-JUN-1996; 96JP-0155575.
XX	
PA	(AJIN) AJINOMOTO KK.
PA	(AJIN) AJINOMOTO CO INC.
XX	
PI	Kawahara Y, Kimura E, Kuwabara Y, Nakamatsu T;
XX	
DR	WPI: 1998-063131/06.
DR	N-PSDB; AAV13157.
XX	
PT	Improved fermentative production of amino acids - by recombina
PT	of gene with modified form introduced on plasmid containing
PT	temperature sensitive replicon
XX	
PS	Example 3; Pages 56-63; 77pp; Japanese.
XX	
CC	The present sequence, which is Brevibacterium lactofermentum
CC	alpha-ketoglutaric acid dehydrogenase, was used in the prepara
CC	of a novel microorganism possessing a gene in a form acting
CC	disadvantageously on the production of a target substance by
CC	fermentation, modified by the introduction of a plasmid contain
CC	a functional version of the gene together with a temperature
CC	sensitivity replicon. The gene is preferably a gene encoding an
CC	enzyme produced by a species capable of biosynthesizing the tar
CC	substance, and the microorganism containing the plasmid is cult
CC	at a temperature allowing the plasmid to replicate, then at a
CC	temperature where plasmid replication does not occur so that th
CC	plasmid is therefore detached from the microbial cell.
CC	The microorganism can be used for the efficient production of a
CC	acids and other target substances by fermentation with improved
CC	yield, in particular, L-glutamic acid production is enhanced us
CC	alpha-ketoglutaric acid dehydrogenase gene, and L-lysine produ
CC	using homoserine dehydrogenase gene.
XX	
SQ	Sequence 1257 AA;

	Query Match	4.7%;	Score 89.5;	DB 19;	Length 1257;
	Best Local Similarity	20.2%;	Pred. No. 5.7;		
	Matches 62;	Conservative 37;	Mismatches 123;	Indels 85;	Gaps 13;
QY	65	RGHVGSSGNDP	SLSPYLIIAILKOLET	VTVVVAABHFYGDKKNYIILTEAMRAVITRLALF	124
		: : : :	: : : : :		
Db	781	rgh-nea	dpsmtgpkmyelitretvraqytdllgrgd----	lsnedaavvr--df	832
QY	125	RNSGYKMLLQGGET	PNEEKSDNQESONRAGNSGNLPGHLGNQ-----		169
		: : : :	: : :		
Db	833	hd-----	qmesvfnevggkkaeqatigtysqklphgletnl	sraellegafan	885
QY	170	-----NHNH-	-----PNWLEGRAMSALSSFQGNARTTTTSSTPGNSR	204	

Db	886	tpgefnyhprvapkrvsvtgegidwagw-ellafgslansgrlvraged-----sr	940
QY	205	R-IQHQAQVIEPPMIKERRRTMSELLTEKGVNGALFAICEVLYITRPLIYVLFIRKYGV	262
Db	941	rgtftqrhavaidpataeefnplhelaqsgnngkf-----lvynsalteyagmgfeygy	995
QY	263	R-----SWIPWASISLSDVTLCMGLL-----ANSKWGEKSKQV-----HFGSGPEKDE	304
Db	996	svgnedsvvvweaafgafangatliideyvsgeakwgtsklilllphgyegqpdhss	1055
QY	305	LRRRLKI 311	
Db	1056	arierfI 1062	
RESULT 7			
AA	388889		
ID	AAAR38889	standard; Protein; 3165 AA.	
XX	AC	AAAR38889;	
DT	25-FEB-1994	(first entry)	
DE	Sequence encoded by ORF B of the long double-stranded cytoplasmic		
DE	RNA (L-dsRNA) present in the hypovirulent strain Ep713.		
XX	KW	Hypovirulent; spore; pathogenic fungus; chestnut blight; papain;	
KW	hypovirulence associated virus; potyvirus; genetic element.		
XX	OS	Cryphonectria (Endothia) parasitica strain Ep713.	
XX	WO9316170-A.		
PD	19-AUG-1993.		
XX	03-FEB-1993;	93WO-US01024.	
XX	06-FEB-1992;	92US-0832117.	
XX	(CHOI/) CHOI G H.		
PA	(NUSS/) NUSS D L.		
XX	Choi GH, Nuss DL;		
XX	WPI; 1993-272875/34.		
DR	N-PSDB; AAQ47380.		
XX	Fungi and fungal spores or components - modified to confer		
PT	transmissible hypo:virulent phenotype, useful for controlling		
PT	fungal diseases, e.g. chestnut blight		
XX	Claim 23; Fig 1; 69pp; English.		
XX	Natural strains of C. parasitica which are hypovirulent contain a		
CC	cytoplasmic determinant that is transferred by hyphal anastomosis.		
CC	The determinant is a ds RNA species which is believed to be of viral		
CC	origin. The large dsRNA (L-dsRNA) present in hypovirulent C.		
CC	parasitica strain Ep713 encodes two large polyproteins (AAR38888,		
CC	AAR38889) that undergo autoproteolytic processing during translation.		
CC	The AA sequences of these polyproteins contain five domains with		
CC	significant similarity to conserved domains within the protein		
CC	products encoded by members of the potyvirus gp. of +ve strand RNA		
CC	plant viruses, and a common ancestry is implied. The term		
CC	hypovirulence-associated virus (HAV) is used to denote this class		
CC	of genetic element. ORF encodes two polypeptides, p29 and p40, that		
CC	are released from a polyprotein, p69, by autocatalysis mediated by		
CC	p29. Cleavage occurs between Gly-248 and Gly-249 during translation		
CC	and is dependent on residues Cys-162 and His-215. Expression of ORF		
CC	B also involves an autoproteolytic event in which a 48 kDa		
CC	polypeptide, designated p48, is released from the N-terminal of the		
CC	encoded polyprotein. Cleavage of p48 occurs between Gly-418 and		
CC	Ala-419 and is dependent upon residues Cys-341 and His-388. Both		

XX	16-JUN-1999	(first entry)	
DT	HIV Tat protein.		
XX	HIV; Tat protein; epitope; human immunodeficiency virus; immune response;		
XX	viral multiplication reduction; chronic viraemia.		
KW	Human immunodeficiency virus type 1.		
XX	WO902185-A1.		
XX	21-JAN-1999.		
XX	10-JUL-1998;	98WO-0514332.	
XX	11-JUL-1997;	97US-0893853.	
XX	(THYM-) THYMON LLC.		
XX	Goldstein G;		
XX	WPI; 1999-120516/10.		
DR	N-PSDB; AAX28268.		
XX	Peptides containing epitopes of human immune deficiency virus Tat		
PT	protein - used to induce antibodies reacting with HIV-1 Tat		
PT	proteins, used to, e.g. reduce HIV-1 levels in both infected and		
PT	uninfected patients		
XX	Claim 20; Page 77-78; 140pp; English.		
XX	This sequence represents the HIV Tat protein, fragments of this sequence		
CC	are claimed for use in the composition of the invention.		
CC	The invention relates to a composition containing a peptide of formulae		
CC	(I)-(IV): R1-VDP-y-LEP-R2 (I); R3-K-x-LGTSIGRKK-R4 (II);		
CC	R5-RR-x'-zay'-S-R6 (III); R7-SQ-x''-HQ-y''-SLSKQP-R8 (IV); where:		
CC	y = R, K, S, N or ornithine; R1, R3, R5 and R7 = hydrogen, lower alkyl or		
CC	alkanoyl, or 1-5 (for R1 and R3) or 1-3 (for R5 and R7) amino acids (aa),		
CC	optionally substituted by lower alkyl or alkanoyl; R2 = free hydroxy,		
CC	amide or 1-4 additional aa, optionally substituted by lower alkyl or		
CC	alkanoyl; x = G or A; R4 = hydroxy, amide or 1-5 additional aa, G or S;		
CC	optionally substituted by amide; x' = A, P, S or Q; y' = D, N, G or S;		
CC	z = P or H; a = Q or P; R6 = hydroxy or amide; x'' = N or T;		
CC	y'' = A or V, and R8 = hydroxy, amide or 1-3 additional aa, optionally		
CC	substituted by amide. The compositions, and synthetic molecules, viruses,		
CC	or bacteria containing them, are used to induce Ab that react with over		
CC	95, best 99, % of known human immunodeficiency virus (HIV)-1 Tat		
CC	proteins, particularly to reduce HIV-1 levels in both infected and		
CC	uninfected subjects (provided they are still immunocompetent). Antibodies		
CC	(Ab) against the peptides reduce viral multiplication during initial		
CC	infection and minimise the chronic viraemia that leads to acquired immune		
CC	deficiency syndrome. Ab are used to reduce HIV-1 levels in infected		
CC	subjects unable to mount an immune response (passive immunisation).		
CC	Peptides (I)-(II) represent four different epitopes in Tat protein, and a		
CC	relatively small number of them will induce antibodies cross-reactive is		
CC	with most strains and variants of Tat. The content of the composition is		
CC	easily adjusted to account for and new strains discovered. They can be		
CC	administered in the early stages of, or before, infection and do not		
CC	promote selection of escape mutants.		
XX	Sequence	302 AA;	
SQ			
XX	Query Match	4.6%; Score 87.5; DB 20; Length 302;	
XX	Best Local Similarity	24.1%; Pred. No. 1;	
XX	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		
QY	124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLGPHGLGN-QN 170		
QY	Best Local Similarity	24.1%; Pred. No. 1;	
QY	Matches	45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;	
QY	64 PRGHVSGSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123		
DB	91 pwehpgssgvdprl-----epwnhl---gssgvdhrlep 121		

Db	237	gsrqrir	243
<hr/>			
Db	295	VHFSGPEKDELRRKKLIWALYLM	317
QY	499	----gaqavaaaqr-llhaialm	516
Db	499	----	
<hr/>			
Db	237	gsrqrir	243
RESULT	13		
ID	AAP93357	standard; protein; 1522 AA.	
XX	AC	AAP93357;	
XX	DT	22-MAR-1991 (first entry)	
DE	DE	Sequence of the catalytic domain of Bordetella pertussis	
KW	KW	adenylate cyclase (AC).	
XX	KW	vaccine; ss.	
XX	OS	Bordetella pertussis.	
XX	PN	FR2618453-A.	
XX	PD	27-JAN-1989.	
XX	PF	24-JUL-1987; 87FR-0010589.	
XX	PR	24-JUL-1987; 87FR-0010589.	
XX	PA	(INSP) INST PASTEUR.	
XX	PI	Danchin A, Glaser P, Ullmann A;	
DR	DR	WFI; 1989-079098/11.	
DR	N-PSDB;	AAN90659.	
XX	PT	Cloning and expressing genes in multi-protein function system - by	
PT	PT	transforming host cell with indicator gene producing protein which	
XX	XX	interacts with expression prod. of the gene being cloned.	
PS	Claim	14; Fig 2(a-f); 24pp; French.	
CC	Sequences	encoding at least a part of AC able to hybridise with a	
CC	gene	expressing a protein with AC activity, and proteins with AC	
CC	activity	able to form immune complexes with specific antibodies are	
CC	claimed.	The isolated AC gene can be used to produce Bordetella	
CC	pertussis	toxin.	
XX	Sequence	1522 AA;	
SQ			
<hr/>			
Query Match	4.5%;	Score 87;	DB 10; Length 1522;
Best Local Similarity	21.7%;	Pred. No. 14;	
Matches	70;	Conservative	39; Mismatches 104; Indels 110; Gaps 17;
QY	25	LTWLLPKFSAISIGPEAVTAF-----LGIFTTINEHIENAPTGRGHV-----	68
Db	274	llwkia ragarsavgtearrifrydgmigvitdf-elevrnalnrrahavgagdvvqh	332
QY	69	GSSGNPDSLSPYLIALILKDLETVEVAEAEHFVGDKKNYIILTAMKAVIRLAFNRSG	128
Db	333	gteqnp---fp-----eadek-----ifvsa-----tg	354
QY	129	YKMLGG---ETPNEEKDSNOSESQNRA-GNSGRNLPGHLGNQHNNHPWNLEGRAMSA	184
Db	355	esqlmrqlkeyiggrgegyvfyenrayvagkslfdgiga-----pgypsgr---	406
QY	185	LSFGONARTTSSTPGWRRTQHQOAVEPMIKERRMTSELITEKGVNALFAIGEV	244
Db	407	-skfsdpvietvapsglr-----pslgaverqdsgydsldvgvsrslgev	454
QY	245	-----LYITRPLIVLFIRKYGVRSWIPIWAISLSVDTLGMGLLANSKWGEKSQ	294
Db	455	smaaveaeaelentrvilha-----garq-----ddaepgvsgasahwqgralq	498

PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139753.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143342.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	26-OCT-1999;	99US-0161360.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147192.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			

Query Match

Best Local Similarity 4.5%; Score 86.5; DB 21; Length 197;

Matches 19; Conservative 16; Mismatches 25; Indels 1; Gaps 1;

QY	55	EHIIENAPTRGHVSSGNDPSLSYPLLIATLKDLFTVVEAAEHFYGDKNWYILTEA	114
Db	85	qhlledstlvkvigidgsvklfndygsi-kvdedlsdlnqkigdkkwglsaltet	143
QY	115	M 115	
Db	144	i 144	

RESULT 15
AAG52858
ID AAG52858 standard; Protein; 288 AA.
XX AC AAG52858;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 67237.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 30-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

Search completed: July 11, 2001, 11:35:32
Job time: 31 sec

PR	20-AUG-1999;	99US-0149722.	Query Match	4.5%;	Score 86.5;	DB 21;	Length 288;
PR	20-AUG-1999;	99US-0149723.	Best Local Similarity	31.1%;	Pred No. 1.2;	Mismatches	1; Gaps 1;
PR	20-AUG-1999;	99US-0149929.	Matches	19;	Conservative	16;	
PR	23-AUG-1999;	99US-0149902.					
PR	23-AUG-1999;	99US-0149930.					
PR	25-AUG-1999;	99US-0150566.					
PR	26-AUG-1999;	99US-0150884.					
PR	27-AUG-1999;	99US-0151065.					
PR	27-AUG-1999;	99US-0151066.					
PR	27-AUG-1999;	99US-0151080.					
PR	30-AUG-1999;	99US-0151303.					
PR	31-AUG-1999;	99US-0151438.					
PR	01-SEP-1999;	99US-0151930.					
PR	07-SEP-1999;	99US-0152363.					
PR	10-SEP-1999;	99US-0153070.					
PR	13-SEP-1999;	99US-0153758.					
PR	15-SEP-1999;	99US-0154018.					
PR	16-SEP-1999;	99US-0154039.					
PR	20-SEP-1999;	99US-0154779.					
PR	22-SEP-1999;	99US-0155139.					
PR	23-SEP-1999;	99US-0155486.					
PR	24-SEP-1999;	99US-0155659.					
PR	28-SEP-1999;	99US-0156458.					
PR	29-SEP-1999;	99US-0156596.					
PR	04-OCT-1999;	99US-0157117.					
PR	05-OCT-1999;	99US-0157753.					
PR	06-OCT-1999;	99US-0157865.					
PR	07-OCT-1999;	99US-0158029.					
PR	08-OCT-1999;	99US-0158232.					
PR	12-OCT-1999;	99US-0158369.					
PR	13-OCT-1999;	99US-0158293.					
PR	13-OCT-1999;	99US-0159294.					
PR	13-OCT-1999;	99US-0159295.					
PR	14-OCT-1999;	99US-0159329.					
PR	14-OCT-1999;	99US-0159330.					
PR	14-OCT-1999;	99US-0159331.					
PR	14-OCT-1999;	99US-0159637.					
PR	14-OCT-1999;	99US-0159636.					
PR	18-OCT-1999;	99US-0159584.					
PR	21-OCT-1999;	99US-0160741.					
PR	21-OCT-1999;	99US-0160767.					
PR	21-OCT-1999;	99US-0160768.					
PR	21-OCT-1999;	99US-0160770.					
PR	21-OCT-1999;	99US-0160814.					
PR	21-OCT-1999;	99US-0160815.					
PR	22-OCT-1999;	99US-0160980.					
PR	22-OCT-1999;	99US-0160981.					
PR	22-OCT-1999;	99US-0160989.					
PR	25-OCT-1999;	99US-0161404.					
PR	25-OCT-1999;	99US-0161405.					
PR	25-OCT-1999;	99US-0161406.					
PR	26-OCT-1999;	99US-0161359.					
PR	26-OCT-1999;	99US-0161360.					
PR	26-OCT-1999;	99US-0161361.					
PR	28-OCT-1999;	99US-0161920.					
PR	28-OCT-1999;	99US-0161992.					
PR	28-OCT-1999;	99US-0161993.					
PR	29-OCT-1999;	99US-0162142.					
QY	55	EHIEIENAPTRGRHVGSSGNDPSLPLIAILKDLFTVVVEAAEHFYGDKKWNYIILTEA	114				
Db	176	qhlledstlvkvvgigidgsvklfhdvgvsi-kdvedlsdianqkggdkkgwlasltet	234				
QY	115	M 115					
Db	235	I 235					

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: July 12, 2001, 00:49:05 ; Search time 119.83 Seconds
(without alignments)
7770.826 Million cell updates/sec

Title: US-09-545-072a-1
Perfect score: 1483
Sequence: 1 attgcaaccaggagagaaaa.....attataaaaaaaaaaaaaa 1483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1483	100.0	1483	21	Arabidopsis thalia
2	85.4	5.8	936	22	Oligonucleotide D1
3	85.4	5.8	936	22	Oligonucleotide D1
4	85.4	5.8	936	22	Oligonucleotide D1
5	85.4	5.8	936	22	Oligonucleotide D1
6	85.4	5.8	936	22	Oligonucleotide D1
7	85.4	5.8	936	22	Oligonucleotide D1
8	83.8	5.7	936	22	Oligonucleotide D1
9	83.8	5.7	936	22	Oligonucleotide D1
10	83.8	5.7	936	22	Oligonucleotide D1
11	83.8	5.7	936	22	Oligonucleotide D1

12	83.8	5.7	936	22	AAF58262	Oligonucleotide D2
13	83.8	5.7	936	22	AAF58255	Oligonucleotide D1
14	51.2	3.5	244	22	AAF58238	Oligonucleotide D1
15	45.6	3.1	244	22	AAF58238	Oligonucleotide D1
16	42	2.8	2070	21	AAC76167	Human ORFX ORF1722
17	41.6	2.8	2123	18	AAT59701	PTH-like peptide D
18	40.4	2.7	603	21	AAA16288	Human colon cancer
19	40.2	2.7	199	22	AAF17640	Human breast cancer
20	40.2	2.7	887	21	AAC59297	Human secreted pro
21	39.6	2.7	1361	21	AAA26399	Human secreted pro
22	39.6	2.7	1547	21	AAC59424	Human secreted pro
23	39.6	2.7	1738	21	AAC99089	Human pancreatic c
24	39.4	2.7	441	20	AAZ09474	pl35-NT3 construct
25	38.4	2.7	1528	18	AAT90733	Human consensus b1
26	39.4	2.7	4120	20	AAZ09473	Human RNA helicase
27	39	2.6	396	22	AAF94827	Human ovarian canc
28	39	2.6	721	21	AAC73912	Human secreted pro
29	39	2.6	2833	21	AAC77810	Human cancer assoc
30	39	2.6	3021	21	AAC75886	Human ORFX ORF1441
31	38.8	2.6	1495	21	AAZ7985	Corn CCR4 transcri
32	38.8	2.6	2364	21	AAZ33323	Human secreted pro
33	38.4	2.6	225	21	AAC98734	Human colon cancer
34	38.4	2.6	302	21	AAC98700	Human colon cancer
35	38.4	2.6	775	19	AAV44861	Clone BV66.1 codin
36	38.4	2.6	1671	21	AAC59315	Human secreted pro
37	38.2	2.6	2670	21	AAF18202	Lung cancer associ
38	38	2.6	3047	21	AAZ07586	Drosophila piwi ge
39	38	2.6	7286	20	AAZ57224	WO 9923223 Seq ID
40	38	2.6	7938	20	AAZ57235	WO 9923223 Seq ID
41	37.8	2.5	818	20	AAZ16631	Human gene express
42	37.8	2.5	818	20	AAZ16629	Human gene express
43	37.8	2.5	1087	21	AAC32912	Arabidopsis thalia
44	37.8	2.5	1437	21	AAC59567	Human secreted pro
45	37.6	2.5	248	21	AAC95297	Cat flea head and

ALIGNMENTS

RESULT	1
AAA88782	
ID	AAA88782 standard; cDNA; 1483 BP.
XX	
AC	AAA88782;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Arabidopsis thaliana SSE1 cDNA.
XX	
KW	SSE1; shrunken seed gene; storage reserve; storage protein;
KW	oil body; transgenic plant; ss.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	122..1225
XX	/*tag= a
PN	WO200061735-A1.
XX	
PD	19-OCT-2000.
XX	
PF	07-APR-2000; 2000WO-US09192.
XX	
PR	08-APR-1999; 99US-0128651.
XX	
PA	(GEHO) GEN HOSPITAL CORP.
XX	
PI	Lin Y;
XX	
DR	WPI; 2000-679483/66.
DR	P-ESDB; AAB19718.
XX	

PT Novel shrunken seed gene useful for producing transgenic plants having
PT altered production of food storage reserve material, intracellular
PT transport of storage protein and formation of protein or oil bodies -
XX
PS Claim 8; Page 57; 64pp; English.

The present sequence is that of *Arabidopsis thaliana* SSEL (shrunk seed) cDNA, which was isolated from a seedling cDNA library using a partial genomic clone as probe. SSEL encodes a protein (see AAB19718) that, when expressed in a cell of a plant, modifies or alters the production of a food storage reserve material (e.g. protein, lipid or carbohydrate storage reserve), facilitates the intracellular transport of a storage protein, or facilitates the formation of protein or oil bodies. The invention provides a transgenic plant (or plant cell, plant tissue, plant organ or plant component) which includes a recombinant SSEL transgene that modifies the production of food storage reserves, thereby increasing nutritional value. An antisense construct is useful for modifying desiccation tolerance.

Sequence 1483 BP; 461 A; 268 C; 347 G; 407 T; 0 other;

Query Match 100.0%; Score 1483; DB 21; Length 1483;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	attgcaaccggaagagaagaaaaatcacagagattgatttaacgtgaatgaattttgttg	60
Db	1	attgcaaccggaagagaagaaaaatcacagagattgatttaacgtgaatgaattttgttg	60
Qy	61	tttcccaaatcttctcagagaaatagcaagattcaagtttctcagtttctctatctgaagctc	120
Db	61	tttcccaaatcttctcagagaaatagcaagattcaagtttctcagtttctctatctgaagctc	120
Qy	121	aatggaagcttataagcaatgggttttgagagaaatagagagatgatacaatcctttggatc	180
Db	121	aatggaagcttataagcaatgggttttgagagaaatagagagatgatacaatcctttggatc	180
Qy	181	cttgccaaacggattgacatggctgctctcagagaagtttctcgtctcagagattggacc	240
Db	181	cttgccaaacggattgacatggctgctctcagagaagtttctcgtctcagagattggacc	240
Qy	241	agaagcagtgtaacggcttttttggcgcatattcaacaacgataataatgaaacataatgaaaa	300
Db	241	agaagcagtgtaacggcttttttggcgcatattcaacaacgataataatgaaacataatgaaaa	300
Qy	301	tgtctccaaacacctgtggccatgttgatcttcgggaatgataccatcccttcttatcc	360
Db	301	tgtctccaaacacctgtggccatgttgatcttcgggaatgataccatcccttcttatcc	360
Qy	361	actactcatcgccatcctcaagagatttggaaactgtttgtggaagtggcagctgaacactt	420
Db	361	actactcatcgccatcctcaagagatttggaaactgtttgtggaagtggcagctgaacactt	420
Qy	421	ctatggagacaaaaatggaactacatttctcactgaagctatgaaaggtgtctattag	480
Db	421	ctatggagacaaaaatggaactacatttctcactgaagctatgaaaggtgtctattag	480
Qy	481	gttagccttgttcgggaataatagtggtgataagatgtcttctcaaggagggggaaacacctaa	540
Db	481	gttagccttgttcgggaataatagtggtgataagatgtcttctcaaggagggggaaacacctaa	540
Qy	541	tgaggagaagaagattctaaccaatccgagtcgcgaaaaatagagctggtaattcgggtagaaa	600
Db	541	tgaggagaagaagattctaaccaatccgagtcgcgaaaaatagagctggtaattcgggtagaaa	600
Qy	601	tctcgggacctaatggtctttggaaaccaaataatcataatccatggaactttgggaagcag	660
Db	601	tctcgggacctaatggtctttggaaaccaaataatcataatccatggaactttgggaagcag	660
Qy	661	ggcgatgtctgctttaaagttcattttggtcagaatgcaagaacaaacacatacttcttcccc	720

Db	661	ggcgatgctgctttaaagtttcatttggtcagaatgcagaacaacaacatcttctacccc	720
Qy	721	cggttgggtctcgaagaattcaacatcagcaacgagttatagagctccaatgatcaaggga	780
Db	721	cggttgggtctcgaagaattcaacatcagcaacgagttatagagctccaatgatcaaggga	780
Qy	781	gaagcgaagaacgagatcgagagctactactagagaaggggtttaatggagcgttggttgc	840
Db	781	gaagcgaagaacgagatcgagagctactactagagaaggggtttaatggagcgttggttgc	840
Qy	841	gattggtgaggttctttacataacgagacgcctatttacgttcttttcacagaaaaa	900
Db	841	gattggtgaggttctttacataacgagacgcctatttacgttcttttcacagaaaaa	900
Qy	901	tggagtcgcatcttggtattccctgggctatatcgctttcttggacacactgggagtggg	960
Db	901	tggagtcgcatcttggtattccctgggctatatcgctttcttggacacactgggagtggg	960
Qy	961	tcttcttgcaaaatcgaaagtgtgtgggagagaagacgaagcaagtcacattctcaggacc	1020
Db	961	tcttcttgcaaaatcgaaagtgtgtgggagagaagacgaagtcacattctcaggacc	1020
Qy	1021	tgaaaaggtgagctgagagacgcaaaaactgataTggcattgtacctcatgagagatcc	1080
Db	1021	tgaaaaggtgagctgagagacgcaaaaactgataTggcattgtacctcatgagagatcc	1080
Qy	1081	attcttcaccaagtatacacaaggcagaagctggaaagctctcaaaaagagctggaaactaat	1140
Db	1081	attcttcaccaagtatacacaaggcagaagctggaaagctctcaaaaagagctggaaactaat	1140
Qy	1141	tcaattgatcggattcctcacagagaagagattgtggagcttttggaggagagctcagtcacg	1200
Db	1141	tcaattgatcggattcctcacagagaagagattgtggagcttttggaggagagctcagtcacg	1200
Qy	1201	gtacacttacatcgggatacgTgaggtTaaagcgttttacttatggtttatatgccaacgg	1260
Db	1201	gtacacttacatcgggatacgTgaggtTaaagcgttttacttatggtttatatgccaacgg	1260
Qy	1261	aagaatttgccattgtTggaatgcttttttagatcaatcaaaagcgtctcacagatttctt	1320
Db	1261	aagaatttgccattgtTggaatgcttttttagatcaatcaaaagcgtctcacagatttctt	1320
Qy	1321	agggaaatggtttcaggccttttgtTagaatttgtttattgcaacaggtagagaacataa	1380
Db	1321	agggaaatggtttcaggccttttgtTagaatttgtttattgcaacaggtagagaacataa	1380
Qy	1381	ccatagacagatgactcaagagataaagcttctctatgtctcaagaagaattggacgatac	1440
Db	1381	ccatagacagatgactcaagagataaagcttctctatgtctcaagaagaattggacgatac	1440
Qy	1441	gaataaaacaagcatcatTaaagattTaaaaaataaaaaaataaa	1483
Db	1441	gaataaaacaagcatcatTaaagattTaaaaaataaaaaaataaa	1483

RESULT 2

RESUL Z
AAF58252/C

ID AAF58252 standard; DNA; 936 BP.

XX
XXAA
AC AAF58252:

AC
XX
AAAF 30232;

2A-APP-2001 / first en

DT 24-APR-2001 (FIRST ENCLY)
YY

XX
DE
Oligonucleotide D193DE
v v
OLIGONUCLEXX
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543

KW Electron-tra

33

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS
XX Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX

```
Query Match      5.8%; Score 85.4; DB 22; Length 936;
Best Local Similarity 0.7%; Pred. No. 2.4e-13;
Matches 5; Conservative 440; Mismatches 306; Indels 0; Gaps 0;

QY 733 aagaattcaacacagcagcattatagagctccaatgatcaagagagcggaagaac 792
DB 762 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 703

QY 793 gatgtcgcagctactactagagaggtgttaatgagcgtgtttgcgattggtgaggt 852
DB 702 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 643

QY 853 tcttacaacagagacgcctcattacgttcttctcagacaaatattgagtcgagtc 912
DB 642 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 583

QY 913 ttggattccttggtgctatgcgttcttcttgacacactgggagtggtctcttgcgaa 972
DB 582 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 523

QY 973 ttcgaagtgtggggagagagcaagcaagtcattctcagagacccgaaagatga 1032
DB 522 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 463

QY 1033 gctgagagacgaaactgatgtggcattgacctcagagagatccattctcaccac 1092
DB 462 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 403

QY 1093 gtacacaaggcagaagctggaagctctcaaaagagcggaactaatcttcattgacgg 1152
DB 402 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 343

QY 1153 attctcacagagaagattgtgagcgttttggagggagcgcagtcacggtacactacat 1212
DB 342 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 283

QY 1213 atcgggatcgtgaggttaagcgttttactatggtttatgatcgaacggaagattgccc 1272
DB 282 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 223

QY 1273 attgttggaatgcttttttagatcatcaaaaggctcctacagatttcttagggaatggtt 1332
```

```
RESULT 7
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
XX
KW gene expression; ss.
XX
OS Synthetic.
```

```
XX WO200107665-A2.
XX
XX 01-FEB-2001.
PD
XX
XX 26-JUL-2000; 2000WO-US20476.
PF
XX
XX 26-JUL-1999; 99US-0145695.
PR
XX 17-MAR-2000; 2000US-0190259.
PR
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA
XX Umek RM;
PI
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SQ
```

```
Query Match      5.8%; Score 85.4; DB 22; Length 938;
Best Local Similarity 0.7%; Pred. No. 2.4e-13;
Matches 5; Conservative 440; Mismatches 306; Indels 0; Gaps 0;

QY 733 aagaattcaacacagcagcattatagagctccaatgatcaagagagcggaagaac 792
DB 762 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 703

QY 793 gatgtcgcagctactactagagaggtgttaatgagcgtgtttgcgattggtgaggt 852
DB 702 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 643

QY 853 tcttacaacagagacgcctcattacgttcttctcagacaaatattgagtcgagtc 912
DB 642 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 583

QY 913 ttggattccttggtgctatgcgttcttcttgacacactgggagtggtctcttgcgaa 972
DB 582 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 523

QY 973 ttcgaagtgtggggagagagcaagcaagtcattctcagagacccgaaagatga 1032
DB 522 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 463

QY 1033 gctgagagacgaaactgatgtggcattgacctcagagagatccattctcaccac 1092
DB 462 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 403

QY 1093 gtacacaaggcagaagctggaagctctcaaaagagcggaactaatcttcattgacgg 1152
DB 402 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 343

QY 1153 attctcacagagaagattgtgagcgttttggagggagcgcagtcacggtacactacat 1212
DB 342 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 283

QY 1213 atcgggatcgtgaggttaagcgttttactatggtttatgatcgaacggaagattgccc 1272
DB 282 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 223

QY 1273 attgttggaatgcttttttagatcatcaaaaggctcctacagatttcttagggaatggtt 1332
```


KW Electron-transfer group; ETM; mismatch; genotyping;
 XX gene expression; ss.
 OS Synthetic.
 XX WO200107665-A2.
 XX 01-FEB-2001.
 XX 26-JUL-2000; 2000WO-US20476.
 XX 26-JUL-1999; 99US-0145695.
 XX 17-MAR-2000; 2000US-0190259.
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 XX Umek RM;
 XX WPI; 2001-159728/16.
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface -
 XX
 XX Example 6; Page 128; 159pp; English.
 XX The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX
 XX Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
 SQ

Db 501 560
 QY 1273 attgttggaatgcttttttagatcatcaaaagcctcctacagattcttagggaatggttt 1332
 XX
 Db 561 620
 QY 1333 caggcttttttagaaattgtttattgcaacaggttagagaacataaccatacacagat 1392
 XX
 Db 621 680
 QY 1393 gtatctgaagagataaagcttctctatgtctaaagaaatggaccgatacagaataaaacaag 1452
 XX
 Db 681 740
 QY 1453 catcattaaagattaaaaaaataaaaaaa 1483
 XX
 Db 741 771

RESULT 13
 AAF58255
 ID AAF58255 standard; DNA; 938 BP.
 XX
 AC AAF58255;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Oligonucleotide D1876.
 XX
 KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 XX
 PN WO200107665-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-US20476.
 XX
 PR 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI Umek RM;
 XX
 DR WPI; 2001-159728/16.
 XX
 PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface -
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX
 SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 5.7%; Score 83.8; DB 22; Length 936;
 Best Local Similarity 0.5%; Pred. No. 6.6e-13;
 Matches 4; Conservative 440; Mismatches 307; Indels 0; Gaps 0;

QY 733 aagaattcaacatcagcaagcagttatagagctccaatgatcaagagagcgagaac 792
 Db 21 80
 QY 793 gatgcgagctactactagagaggtgttaatggagcgtttgttcgcttggtgaggt 852
 Db 81 140
 QY 853 tctttacataacagagacgcgtcatttacgttcttttcatcagaaaaatagtgatcgatc 912
 Db 141 200
 QY 913 ttggattccttggtctatagcttctgtggacacactgggagtggtcttctgcaa 972
 Db 201 260
 QY 973 ttcgaagtgtgggagagagcaagcagctcatttctcagacctgaaaggatga 1032
 Db 261 320
 QY 1033 gctgaggagacgaaactgatattggtcattgtacctcatgagagatccatttcacaa 1092
 Db 321 380
 QY 1093 gtacacagcgagaagctggaagctctcaaaagaagctggaactaattcattgacgg 1152
 Db 381 440
 QY 1153 attctcacagagaagattgtggagctttttgagggagctcagtcacggtacattacat 1212
 Db 441 500
 QY 1213 atcgggacgtgaggttaagcgttttacttatggttttatgcaacgagaatattgcc 1272

Query Match 5.7%; Score 83.8; DB 22; Length 938;
 Best Local Similarity 0.5%; Pred. No. 6.6e-13;
 Matches 4; Conservative 440; Mismatches 307; Indels 0; Gaps 0;

QY 733 aagaattcaacatcagcaagcagttatagagctccaatgatcaagagagcgagaac 792
 Db 21 80

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 00:18:40 ; Search time 96.91 Seconds
(without alignments)
2835.013 Million cell updates/sec

Title: US-09-545-072A-1
Perfect score: 1483
Sequence: 1 attgcaaccaggaagagaaa.....attataaaaaaaaaaaaaa 1483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCRUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38.8	2.6	7218	1	US-08-232-463-14
2	38	2.6	7286	4	US-09-331-581-3
3	38	2.6	7938	4	US-09-331-581-14
4	37.2	2.5	1117	4	US-09-247-373B-33
5	37.2	2.5	3871	2	US-08-599-455B-3
6	36.8	2.5	2837	2	US-08-993-228-11
7	36.4	2.5	2547	3	US-08-262-220-7
8	36.4	2.5	2547	3	US-08-471-733-7
9	36.4	2.5	2547	3	US-08-468-878-7
10	36.4	2.5	2547	3	US-08-750-494-7
11	36	2.4	940	2	US-08-471-717-1
12	36	2.4	2433	1	US-08-136-743B-40
13	36	2.4	2663	1	US-08-136-743B-3
14	35.6	2.4	8920	2	US-08-446-855A-1
15	35.6	2.4	8920	4	US-09-150-741-1
16	35.4	2.4	240	1	US-08-628-417-6
17	35.4	2.4	4137	3	US-09-221-235-1
18	35.4	2.4	4137	4	US-09-221-928-1
19	35.4	2.4	4137	4	US-09-221-527-1
20	35.4	2.4	4137	4	US-09-221-236-1
21	35.4	2.4	4137	4	US-09-221-416-1
22	35.4	2.4	4137	4	US-09-221-245-1
23	35.4	2.4	4137	4	US-09-163-115-1
24	35.4	2.4	4137	4	US-09-221-528-1
25	35.4	2.4	4137	4	US-09-593-553-1
26	35.4	2.4	4137	4	US-09-221-237-1
27	35	2.4	1493	1	US-08-340-820-24

28	35	2.4	1493	1	US-08-593-535-24
29	35	2.4	1582	3	US-08-545-196B-10
30	35	2.4	1582	3	US-08-545-198B-12
31	35	2.4	5173	1	US-08-242-677-1
32	35	2.4	6671	1	US-08-280-443-1
33	35	2.4	6671	1	US-08-457-459-1
34	35	2.4	6671	1	US-08-555-678-1
35	35	2.4	6671	5	PCN-US95-02275-1
C 36	35	2.4	72928	3	US-09-009-913-1
37	34.8	2.3	1882	4	US-09-370-253-1
38	34.6	2.3	1296	1	US-08-385-191A-1
39	34.6	2.3	1441	4	US-08-821-994-63
40	34.4	2.3	746	4	US-09-013-810-1
41	34.4	2.3	1525	4	US-09-157-603-4
42	34.4	2.3	1525	4	US-09-587-436-4
43	34.2	2.3	144	1	US-08-702-344-26
C 44	34.2	2.3	2422	1	US-07-867-106-5
C 45	34.2	2.3	3138	1	US-07-867-106-4

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

[illegible]

```

RESULT      2
US-09-331-581-3
; Sequence 3, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, YUKO
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-OPCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7286
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3

```

```

RESULT      3
US-09-331-581-14
; Sequence 14, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HANA, yuko
; APPLICANT: KUNAGAI, Hiromicho
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-0PCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 7938
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
; US-09-331-581-14

```

RESULT 4
 US-09-247-373B-33
 ; Sequence 33, Application US/09247373B
 ; Patent No. 6168954
 ; GENERAL INFORMATION:
 ; APPLICANT: MCGONIGLE, BRIAN
 ; APPLICANT: O'KEEFE, DANIEL
 ; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
 ; FILE REFERENCE: CL-1108-A
 ; CURRENT APPLICATION NUMBER: US/09/247,373B
 ; CURRENT FILING DATE: 1999-02-10
 ; PRIOR APPLICATION NUMBER: 08/924,747
 ; PRIOR FILING DATE: 1997-09-05
 ; NUMBER OF SEQ ID NOS: 56
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 33
 ; LENGTH: 1117
 ; TYPE: DNA
 ; ORGANISM: SOYBEAN
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1101)
 ; OTHER INFORMATION: M-A OR C
 ; NAME/KEY: unsure
 ; LOCATION: (1104)
 ; OTHER INFORMATION: M-A OR C
 ; NAME/KEY: unsure
 ; LOCATION: (1116)
 ; OTHER INFORMATION: N-G or A or T or C

NAME/KEY: sig_peptide
LOCATION: 380..442
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 443..2242
US-08-471-733-7

Query Match 2.58; Score 36.4; DB 3; Length 2547;
Best Local Similarity 47.4%; Pred. No. 0.47;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
Qy 198 catggtcttcctgagaagtttcttcagagattgacacagcagtaacgctt 257
Db 1913 CTTGTGCTCCGAACTTTCGCGCTTGATTTGTCATCAGCAGCTGAGCTTGTG 1854
Qy 258 ttttgggcattatcacacgataaataatgaaatgtaaaatgctccacacccctg 317
Db 1853 TTGTGGGATAGTTTCTACAGATTTTGTCTTAAATATCATAAAGACCGATAGATGTTA 1794
Qy 318 gcaatgttgatcttcgggaatgatccatcccttcttccactactatcgccatcc 377
Db 1793 TCCATCTAGTCCGCTTCAAGTCTTTTGTGTTGTTCTTATAGCTTGAGATTCAACAT 1734
Qy 378 tcaaggatttggaaactgttggaactggcagctgaacacttctatgga 427
Db 1733 TTAAGCTTGAATGCAATTTTCGGATAGTGTTTTAAATCTTTTCTGTGA 1684

RESULT 9
US-08-468-878-7/c
; Sequence 7, Application US/08468878
; Patent No. 6090586
; GENERAL INFORMATION:
; APPLICANT: BERGSTROM SVEN
; APPLICANT: BARBOUR ALAN G.
; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 SEVENTH STREET, N.W.
; CITY: WASHINGTON
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,878
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,220
; FILING DATE: 20-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: BERGSTROM-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Borrelia garinii

STRAIN: Ip90
IMMEDIATE SOURCE:
CLONE: PJB-101
FEATURE:
NAME/KEY: CDS
LOCATION: 380..2245
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 380..442
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 443..2242
US-08-468-878-7

Query Match 2.58; Score 36.4; DB 3; Length 2547;
Best Local Similarity 47.4%; Pred. No. 0.47;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
Qy 198 catggtcttcctgagaagtttcttcagagattgacacagcagtaacgctt 257
Db 1913 CTTGTGCTTCCTGAACCTTTCGCGCTTGATTTGTCATCAGCAGCTGAGCTTGTG 1854
Qy 258 ttttgggcattatcacacgataaataatgaaatgtaaaatgctccacacccctg 317
Db 1853 TTGTGGGATAGTTTCTACAGATTTTGTCTTAAATATCATAAAGACCGATAGATGTTA 1794
Qy 318 gcaatgttgatcttcgggaatgatccatcccttcttccactactatcgccatcc 377
Db 1793 TCCATCTAGTCCGCTTCAAGTCTTTTGTGTTGTTCTTATAGCTTGAGATTCAACAT 1734
Qy 378 tcaaggatttggaaactgttggaactggcagctgaacacttctatgga 427
Db 1733 TTAAGCTTGAATGCAATTTTCGGATAGTGTTTTAAATCTTTTCTGTGA 1684

RESULT 10
US-08-750-494-7/c
; Sequence 7, Application US/08750494
; Patent No. 6204018
; GENERAL INFORMATION:
; APPLICANT: BERGSTROM SVEN
; APPLICANT: BARBOUR ALAN G.
; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 SEVENTH STREET, N.W.
; CITY: WASHINGTON
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,494
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,220
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: BERGSTROM-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 2547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Borrelia garinii
; STRAIN: Ip90
; IMMEDIATE SOURCE:
; CLONE: pJB-101
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 380...2245
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 380...442
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 443...2242
; US-08-750-494-7

Query Match          2.5%; Score 36.4; DB 4; Length 2547;
Best Local Similarity 47.4%; Pred. No. 0.47;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 198 catggctgctctcagaaagtttctgctcagagattggaccagaagcagtaacggctt 257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1913 CTGTGTGCTCTCGACACITGTCGGCGTGTGATGTTTCATCAGCAGCAGTGAGCTGTG 1854

QY 258 ttttggcatattcacacagataaaatgaacacataattgaaatgctccacacctcgtg 317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1853 TTGTGGGATAGTTTCTTACAGATTTTGTCTTAAATATATATAAGACCGATAGATGTTA 1794

QY 318 gccatgttgattccgggaatgccatcccttcttattccactactatccatgccatcc 377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1793 TCCATCTCTAGTCGGTTCATAGTCCCTTTTGTGCTTCATAGCTTGAGATTTCAACAT 1734

QY 378 tcaagattggaaactgttgtggaagtggcagctgaacacttctatgga 427
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1733 TTAAGCTTTGAATGATTTTCGGATATGGTTTAACTCTTTTCTGTA 1684

RESULT 11
US-08-471-717-1
; Sequence 1, Application US/08471717
; Patent No. 5859337
; GENERAL INFORMATION:
; APPLICANT: Gasser, Charles S.
; APPLICANT: Lippuner, Veronica
; TITLE OF INVENTION: GENES CONFERRING SALT TOLERANCE AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,717
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 2307E-606
```

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 940 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..738
; US-08-471-717-1

Query Match          2.4%; Score 36; DB 2; Length 940;
Best Local Similarity 50.6%; Pred. No. 0.36;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1312 agatttctagggaatggtttcaggctttttagaaattgttattgcaacaggttag 1371
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 759 AGATTTCGTTGTATACGTGTTGAGAGTGTCTAGGAATTTGTTGACTGTACTACCAAT 818

QY 1372 agaacaataaccatagacagatgtatctgaagagataagcttctctatgtctaaagaaatg 1431
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 819 TGGACTTTTGACTGATTCCAATTCCTCTGTTCTTTCAATTTTAAAAAATATTAAACCGATT 878

QY 1432 gaccgatacgataaacaacagcatcattaaagattaaaaaataaaaaa 1483
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 879 CTTTACCACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 930

RESULT 12
US-08-136-743B-40
; Sequence 40, Application US/08136743B
; Patent No. 5459063
; GENERAL INFORMATION:
; APPLICANT: Barry S. Cooperman, Harvey Rubin,
; APPLICANT: Jerome Salem, and Alison L. Fisher
; TITLE OF INVENTION: "Plasmodium falciparum Ribonu-
; TITLE OF INVENTION: cleotide Reductase, DNA Sequences Therefor and Peptide Inhi-
; TITLE OF INVENTION: Thereof"
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The University of Pennsylvania
; STREET: Suite 330
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19104-3246
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,743B
; FILING DATE: 10/14/93
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 3957-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8363
; TELEFAX: (215) 568-5549
; TELEX: No. 5459063e
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2433 nucleotides
```


Qy 1314 atttcttaqqgaatqggtttcagqgccttttqgttagaaaattqgtttattqcaacagqtaqg 1373

Qy 1314 atttcttaqqgaatqggtttcagqgccttttqgttagaaaattqgtttattqcaacagqtaqg 1373

Qy 1314 atttcttaqqgaatqggtttcagqgccttttqgttagaaaattqgtttattqcaacagqtaqg 1373

0' 1338 + + + + + a a a a + + + + + t t t t t a a a c a c a g g t a n a n a a c a t a a c c a t a g a c a g a t a t a t c 1397

Db 281 ATAAATATAAGAAAAA 256

Db 281 ATAAATATAAGAAAAA 256

RESULT 15
 US-09-150-741-1/c
 ; Sequence 1, Application US/09150741
 ; Patent No. 6183996
 ; GENERAL INFORMATION:
 ; APPLICANT: Stewart et al.
 ; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
 ; Patent No. 6183996
 ; TITLE OF INVENTION: Synthetase II
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/09/150,741
 ; CURRENT FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PL6380
 ; EARLIER FILING DATE: 1992-12-16
 ; EARLIER APPLICATION NUMBER: AU93/00617
 ; EARLIER FILING DATE: 1993-12-02
 ; EARLIER APPLICATION NUMBER: 08/446,855
 ; EARLIER FILING DATE: 1995-07-06
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 8920
 ; TYPE: DNA
 ; ORGANISM: Plasmodium falciparum
 US-09-150-741-1

Query Match 2.4%; Score 35.6; DB 4; Length 8920;
 Best Local Similarity 52.7%; Pred. No. 1.6;
 Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
 Qy 1338 ttgttagaattgtgttattgcaacaggtagagaaacataaccatagacagatgtatc 1397
 Db 401 TATATTAATAAGTTTATATTTTCAAGTAATTTATAACAAATGAACACACAAACATATA 342
 Qy 1398 tgaagagataagcttctctatgtctaaagaaatggaccgatacgaataaaacaagcatca 1457
 Db 341 TATATATATATATATATATATATATATATATATATAATAAATAACTTAAATGTATTGTTA 282
 Qy 1458 ttaagattaaaaaiaaaaaaa 1483
 Db 281 ATAAATATAAGAAAAA 256

Search completed: July 12, 2001, 01:48:19
 Job time: 5379 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 00:12:55 ; Search time 1368.44 Seconds
(without alignments)
10244.206 Million cell updates/sec

Title: US-09-545-072A-1
Perfect score: 1483
Sequence: 1 attgcaaccaggagagaaaa.....attaaaaaaaaaaaaaaa 1483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est33.*
26: gb_est34.*
27: gb_est35.*
28: gb_est36.*
29: gb_est37.*
30: gb_est38.*
31: gb_est39.*
32: gb_est40.*
33: em_estba.*
34: em_estfun.*
35: em_esthum1.*
36: em_esthum2.*
37: em_esthum3.*
38: em_esthum4.*
39: em_esthum5.*
40: em_esthum6.*
41: em_esthum7.*
42: em_esthum8.*
43: em_esthum9.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estom1.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estrol.*
83: em_estrol2.*
84: em_estrol3.*
85: em_estrol4.*
86: em_estrol5.*
87: em_estrol6.*
88: em_estrol7.*
89: em_estrol8.*
90: em_estrol9.*
91: em_estrol10.*
92: em_estrol11.*
93: em_estrol12.*
94: em_estrol13.*
95: em_estrol14.*
96: em_estrol15.*
97: em_estrol16.*
98: em_estrol17.*
99: em_estrol18.*
100: em_estrol19.*
101: em_estrol20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est41.*
111: gb_est42.*
112: gb_est43.*
113: gb_est44.*
114: gb_est45.*
115: gb_est46.*
116: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Db	128	TTCTGGTATGACCCCTTGTTGCTCTTCCTGGCTGCTATTGATTGCATANGAACAGTATTC	187
QY	966	ttgca-----aattgaagtgtggggagagaagaagcaagtcacattctcagagac	1019
Db	188	TTTCACATCAATACATGTCAGTGGCTGGTGGGAAGGACCGGAATGTTTCATCTGTCTGCC	247
QY	1020	ctgaaaaggatgagctgagagacgaaactgatatggcattgtacctcatgagagatc	1079
Db	248	TAGAAAGGATGAGGTTAACCGCGCAAGCTGCTATTGTTCTTTTACATATGAGAGATC	307
QY	1080	cattcttcacaaagtacacagaagcagaagctggaaaactctcacaagaagctggaactaa	1139
Db	308	CATTTTTCAGCAAGTATACATAGGCAAAAGACTTGAAGACGAGAGAAAGTTTGGAGCCTA	367
QY	1140	ttccattgatcgattctccacagagaagattgtggagcttttggaggagagctcaagtcac	1199
Db	368	TTCTGTCTATAGGATTCTCACACAAACTGTGTAACTATAAATTTGGAGCTCAGACAC	427
QY	1200	ggtacattcatatcgggatcgtga	1225
Db	428	GATACACTTACATGTCAGGATCGTGA	453
RESULT	5		
BF258527			
LOCUS			
DEFINITION	BF258527	598 bp mRNA	EST 23-FEB-2001
		HVSMF0015P11f Hordeum vulgare seedling root EST library HVCDNA0007	
		(etiolated and unstressed) Hordeum vulgare cDNA clone	
ACCESSION	HVSMF0015P11f		
VERSION			
KEYWORDS			
SOURCE	BF258527		
ORGANISM	BF258527.2	GI:13119474	
		barley.	
REFERENCE			
AUTHORS			
		Hordeum vulgare	
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae	
		; Triticeae; Hordeum.	
		1 (bases 1 to 598)	
		Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu	
		X., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, W., Rambo	
		T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and	
		Wood, J.	
TITLE		Development of a genetically and physically anchored EST resource	
JOURNAL		for barley genomics	
COMMENT		Unpublished (2000)	
		On Nov 16, 2000 this sequence version replaced gi:11187640.	
		Contact: Wing RA	
		Clemson University Genomics Institute	
		Clemson University	
		100 Jordan Hall, Clemson, SC 29634, USA	
		Tel: 864 656 7288	
		Fax: 864 656 4293	
		Email: rwing@clemson.edu	
		Seq primer: AATTAACCTCTACTAAAGG	
		High quality sequence stop: 541.	
FEATURES		Location/Qualifiers	
source		1..598	
		/organism="Hordeum vulgare"	
		/cultivar="Morex"	
		/db_xref="taxon:4513"	
		/clone="HVSMEF0015P11f"	
		/clone_lib="Hordeum vulgare seedling root EST library	
		HVCDNA0007 (etiolated and unstressed)"	
		/tissue_type="Seedling root"	
		/lab_host="JUC121"	
BASE COUNT	155 a	131 c	146 g
ORIGIN			166 t
Query Match	8.7%	Score 128.4;	DB 146; Length 598;
Best Local Similarity	55.6%;	Pred. No. 2.7e-24;	
Matches 268;	Conservative	0;	Mismatches 211; Indels 3; Gaps 1;

JOURNAL
COMMENT


```

/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CSTC2J23"
/clone_lib="sprouting eyes/shoots"
/tissue_type="sprouting tuber"
/dev_stage="13-14 weeks post
lab host="SOLR"

```

/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: XhoI; Various sizes of sprouting eyes (2 mm to 15 mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."

BASE COUNT	133 a	106 c	121 g	172 t
ORIGIN				

	Query Match	5.3%	Score 79.2;	DB 173;	Length 532;
	Best Local Similarity	60.1%;	Pred. No.8.e-11;		
	Matches 197;	Conservative	0;	Mismatches 108;	Indels 23; Gaps 3;
QY	119	tcaatggaagccttaatagcgaatgggtttgagaaaatagagagtattacaaatcccttggga	178		
Db	207	TCTATGGAGGCTTACAAGAGATGGGTCCGAAGAATCGTGATTATGTCGCCAATTGAGT	266		
QY	179	tcctttgccaacggattgacatggctctccctgagaaggtttcttgcttcacagagattgga	238		
Db	267	TCTCTAGCAGCTGGAAATGACATGCCTTCTTCCAGAGCGGTTTCTGTCATCACAAATCGGG	326		
QY	239	ccagaagca-----gtaacggcttttttggggcatattcacacgataa	281		
Db	327	CCAGAAGCAGGTTTTATTGAAATTTTGTCTCTCCGATCTCTGGAATGTTACCACAGTAA	386		
QY	282	atgaacacataatgaaaatgctccaacaccttgtggccaatgttgagatcttcggggaatg	341		
Db	387	ATGAGCATATTATAGAGAACAACTCCAA--CTTCGGTGGCATACTAGTTCCTGCAGAGA---	440		
QY	342	atcccaccctcttcttatccactactcatcgccatcctcaaggatttggaaactgtgttgg	401		
Db	441	CFTCGTTTCTCCCTTGTGCATTATATGCCCTTACTTGTTCGAGACTTGGAAACATTAATTG	500		
QY	402	aagtggcagcgtgaacactcttatggga	429		
Db	501	AAGTAGTAGCTGACGAGCTCTATGGCGA	528		

RESULT	14
BG465365	
LOCUS	
DEFINITION	BG465365 482 bp mRNA EST 20-MAR-2001 EM1_75_E04.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION	BG465365
VERSION	BG465365.1
KEYWORDS	GI:13394341
SOURCE	sorghum.
ORGANISM	sorghum bicolor
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 482)
AUTHORS	Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE	An EST database from Sorghum: developing embryos
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

FEATURES
 FOLIA-NO. Location/Qualifiers
 1. 482
 /organism="Sorghum bicolor"
source

```

/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EML)"
/notes="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda zap II; Site-1: xhoI; Site-2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda zap II. Clones to be sequenced were
prepared by mass excision."
125 a 98 c 114 g 145 t
BASE COUNT
ORIGIN

```

BASE COUNT
ORIGIN

	Query Match	5.3%;	Score 78.2;	DB 154;	Length 482;
	Best Local Similarity	66.1%;	Pred. No. 1.5e-10;		
	Matches 113; Conservative	0;	Mismatches 58;	Indels 0;	Gaps 0;
QY	1053 tatgggcattgtaccctcatgagatgccattcctcaccaagtacacaaggcagaactcg				1112
Dbb					
	1 TGTGGGCCCTTTATTTGATCGAGACCCTTCTTGCCATTACCAACAACGCATCTCC				60
QY	1113 aaagctctcaaaagaagctgggaactaatccattgatcggtaccttcacacagaagaattg				1172
Dbb					
	61 AAAAGGCTGAACAGGTCCTGAATCCGGTGCCAATTGATGGCTCCTTACAGGGAAACTTA				120
QY	1173 tgaagcttttggaggagctcagtcacngtgatacttacatatcggaatogt				1223
Dbb					
	121 TAGAGGTACTTGGAGGGGGTTGCAGAAGAATAATACATACATCACGTTTAT				171

RESULT	15
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

	FEATURES	SOURCE
1	Age	1980 Census
2	Sex	"
3	Ethnicity	"
4	Marital status	"
5	Number of children under 18 years old	"
6	Number of children over 18 years old	"
7	Number of persons in household	"
8	Household type (owner-occupied or rental)	"
9	Value of house	"
10	Year of construction	"
11	Distance to nearest school	"
12	Distance to nearest shopping center	"
13	Distance to nearest park	"
14	Distance to nearest highway	"
15	Distance to nearest airport	"
16	Distance to nearest water body	"
17	Distance to nearest public transit station	"
18	Distance to nearest police station	"
19	Distance to nearest fire station	"
20	Distance to nearest hospital	"
21	Distance to nearest library	"
22	Distance to nearest church	"
23	Distance to nearest synagogue	"
24	Distance to nearest mosque	"
25	Distance to nearest cemetery	"
26	Distance to nearest golf course	"
27	Distance to nearest tennis court	"
28	Distance to nearest swimming pool	"
29	Distance to nearest beach	"
30	Distance to nearest ski resort	"
31	Distance to nearest amusement park	"
32	Distance to nearest zoo	"
33	Distance to nearest museum	"
34	Distance to nearest theater	"
35	Distance to nearest restaurant	"
36	Distance to nearest bar	"
37	Distance to nearest nightclub	"
38	Distance to nearest casino	"
39	Distance to nearest gambling establishment	"
40	Distance to nearest liquor store	"
41	Distance to nearest tobacco shop	"
42	Distance to nearest drugstore	"
43	Distance to nearest pharmacy	"
44	Distance to nearest grocery store	"
45	Distance to nearest supermarket	"
46	Distance to nearest convenience store	"
47	Distance to nearest hardware store	"
48	Distance to nearest home improvement store	"
49	Distance to nearest furniture store	"
50	Distance to nearest clothing store	"
51	Distance to nearest shoe store	"
52	Distance to nearest jewelry store	"
53	Distance to nearest electronics store	"
54	Distance to nearest bookstore	"
55	Distance to nearest record store	"
56	Distance to nearest video store	"
57	Distance to nearest toy store	"
58	Distance to nearest sporting goods store	"
59	Distance to nearest pet store	"
60	Distance to nearest florist	"
61	Distance to nearest hair salon	"
62	Distance to nearest beauty parlor	"
63	Distance to nearest dry cleaner	"
64	Distance to nearest laundromat	"
65	Distance to nearest car wash	"
66	Distance to nearest auto repair shop	"
67	Distance to nearest mechanic	"
68	Distance to nearest oil change service	"
69	Distance to nearest tire service	"
70	Distance to nearest car rental agency	"
71	Distance to nearest truck rental agency	"
72	Distance to nearest moving company	"
73	Distance to nearest storage facility	"
74	Distance to nearest self-storage unit	"
75	Distance to nearest warehouse	"
76	Distance to nearest distribution center	"
77	Distance to nearest factory	"
78	Distance to nearest office building	"
79	Distance to nearest government building	"
80	Distance to nearest courthouse	"
81	Distance to nearest city hall	"
82	Distance to nearest police department	"
83	Distance to nearest fire department	"
84	Distance to nearest hospital	"
85	Distance to nearest nursing home	"
86	Distance to nearest mental health clinic	"
87	Distance to nearest substance abuse treatment center	"
88	Distance to nearest addiction treatment center	"
89	Distance to nearest rehabilitation center	"
90	Distance to nearest correctional institution	"
91	Distance to nearest prison	"
92	Distance to nearest jail	"
93	Distance to nearest juvenile detention center	"
94	Distance to nearest halfway house	"
95	Distance to nearest shelter	"
96	Distance to nearest homeless services center	"
97	Distance to nearest food bank	"
98	Distance to nearest soup kitchen	"
99	Distance to nearest community center	"
100	Distance to nearest senior center	"
101	Distance to nearest day care center	"
102	Distance to nearest preschool	"
103	Distance to nearest elementary school	"
104	Distance to nearest middle school	"
105	Distance to nearest high school	"
106	Distance to nearest college	"
107	Distance to nearest university	"
108	Distance to nearest research institution	"
109	Distance to nearest think tank	"
110	Distance to nearest non-profit organization	"
111	Distance to nearest religious organization	"
112	Distance to nearest political organization	"
113	Distance to nearest labor union	"
114	Distance to nearest professional association	"
115	Distance to nearest trade association	"
116	Distance to nearest industry group	"
117	Distance to nearest business council	"
118	Distance to nearest chamber of commerce	"
119	Distance to nearest economic development corporation	"
120	Distance to nearest public works department	"
121	Distance to nearest transportation department	"
122	Distance to nearest parks and recreation department	"
123	Distance to nearest cultural arts commission	"
124	Distance to nearest historical society	"
125	Distance to nearest preservation commission	"
126	Distance to nearest planning commission	"
127	Distance to nearest zoning board	"
128	Distance to nearest land use committee	"
129	Distance to nearest environmental commission	"
130	Distance to nearest conservation commission	"
131	Distance to nearest watershed protection committee	"
132	Distance to nearest natural resources commission	"
133	Distance to nearest open space advisory board	"
134	Distance to nearest trails committee	"
135	Distance to nearest park advisory board	"
136	Distance to nearest recreation advisory board	"
137	Distance to nearest sports and leisure committee	"
138	Distance to nearest youth council	"
139	Distance to nearest senior citizens council	"
140	Distance to nearest neighborhood association	"
141	Distance to nearest homeowners association	"
142	Distance to nearest residents' association	"
143	Distance to nearest tenants' union	"
144		

BASE COUNT	116 a	100 c	108 g	162 t
ORIGIN				

Query Match 4.9%; Score 72.4; DB 173; Length 486;
Best Local Similarity 72.3%; Pred. No. 5.9e-09;

	Matches	94;	Conservative	0;	Mismatches	36;	Indels	0;	Gaps	0;
Qy	119	tcaatggaagccttataagcaaatgggttttgagaaatagagagatgtatgtacaatcctttgga	178							
Db	211	TCTATGGAGGCTTACAAGAGATGGGTCGAAGGAATCGTGATTATGTCGCCCAATTGAGT	270							
Qy	179	tcctttgccaaacgattgacatggctgcttccctgagagatgtttctgttcagagattgga	238							
Db	271	TCTCTAGCCAGTGGGAATGACATGGCTTCTCCAGAGCGGTTTGCTGCATCAGAAATCGGG	330							
Qy	239	ccagaagcag	248							
Db	331	CCAGAAGCAG	340							

Search completed: July 12, 2001, 01:12:04
Job time: 3549 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 13.09 Seconds
(without alignments)
960.410 Million cell updates/sec

Title: US-09-545-072A-2
Perfect score: 1922
Sequence: 1 MEAYKQWVRNREYVQSFSGS.....EKIVELLEGAQSRVITYISGS 367

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186.5	9.7	391	1	PERG_YARLI
2	96.5	5.0	4486	1	DYH9_HUMAN
3	96	5.0	987	1	YD94_METJA
4	93	4.8	1306	1	MSB2_YEAST
5	89	4.6	549	1	DSX_DROME
6	88	4.6	1706	1	CYAA_BORPE
7	86.5	4.5	4466	1	DYHC_ANTCR
8	85.5	4.4	758	1	CHEA_RHIME
9	85.5	4.4	879	1	MCW3_SCHPO
10	85	4.4	248	1	AMPM_BACHD
11	84	4.4	3411	1	POLG_YEFV1
12	84	4.4	3411	1	POLG_YEFV2
13	83.5	4.3	1325	1	Y309_MYCPN
14	83	4.3	524	1	CP72_CATRO
15	83	4.3	1142	1	ENAM_PIG
16	82.5	4.3	2145	1	CYAA_PODAN
17	82	4.3	566	1	YJ19_SYNY3
18	82	4.3	4829	1	BIR6_HUMAN
19	81.5	4.2	997	1	APG1_YEAST
20	81.5	4.2	1250	1	SSD1_YEAST
21	81	4.2	957	1	KP2_KLULA
22	80.5	4.2	248	1	AMPM_BACSU
23	80.5	4.2	467	1	IRF6_HUMAN
24	80.5	4.2	524	1	CK13_YEAST
25	80.5	4.2	752	1	ABC7_HUMAN
26	79.5	4.1	474	1	CISY_EMEI1
27	79.5	4.1	531	1	UL95_HCMVA
28	79.5	4.1	637	1	YHES_ECOLI
29	79.5	4.1	821	1	CTRL_ARATH
30	79.5	4.1	896	1	APCE_SYNY4
31	79.5	4.1	911	1	P120_MOUSE
32	79.5	4.1	1189	1	YJH6_YEAST
33	79.5	4.1	2076	1	FAS1_YARLI

34 79.5 4.1 2442 1 CBP_HUMAN
35 79.5 4.1 4466 1 DYHC_TRIGR
36 79 4.1 380 1 BCAL_SCHPO
37 79 4.1 590 1 OAM_ASCSU
38 79 4.1 598 1 Y427_HUMAN
39 78.5 4.1 1288 1 VACA_HELPJ
40 78.5 4.1 2194 1 SC16_YEAST
41 78.5 4.1 4273 1 PKSM_BACSU
42 78 4.1 176 1 VLPA_MYCHR
43 78 4.1 427 1 YF09_MYCPN
44 78 4.1 435 1 GAT3_XENLA
45 78 4.1 518 1 KR2_HSV11

ALIGNMENTS

RESULT 1
PEXG_YARLI
ID PEXG_YARLI STANDARD; PRT; 391 AA.
AC P78980;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PEROXISOMAL MEMBRANE PROTEIN PEX16 (PEROXIN-16).
GN PEX16.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E122;
RX MEDLINE=97327755; PubMed=9182661;
RA Eitzen G.A., Szillard R.K., Rachubinski R.A.;
RT "Enlarged peroxisomes are present in oleic acid-grown Yarrowia
lipolytica overexpressing the PEX16 gene encoding an intraperoxisomal
peripheral membrane peroxin.";
RL J. Cell Biol. 137:1265-1278(1997).
CC -!- SUBCELLULAR LOCATION: PEROXISOMAL MEMBRANE-ASSOCIATED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U75433; ABA41724.1; -.
KW Peroxisome; Membrane.
SQ SEQUENCE 391 AA; 44479 MW; F328E3C6CFCE655 CRC64;

Query Match 9.7%; Score 186.5; DB 1; Length 391;
Best Local Similarity 20.7%; Pred. No. 8.5e-09;
Matches 86; Conservative 68; Mismatches 143; Indels 119; Gaps 13;

Qy 1 MEAYKQWVRNREYVQSFSGFANGTLWLLPKFSASEIGPEAVTAFGLGFTTINEHIEN 60
Db 20 LDSYDKFLVRNAASTGSIESTLRVSVYLPGRFNDVEIATETLYAVNLGLYHDTIAR 79
Qy 61 APTPRGHVSGNDPSLSYP-----LLIAILKDLFTVVEVAEEH 99
Db 80 A-----VAASPNAAAVYRPSPHNRYTDWFTKRRKGYKASRAVTFVKFGELVAEMVAKK 133
Qy 100 FYGD-KKNYIILTEAMKAVIRLALFRNSGYKMLLQGETPNEEK-----143
Db 134 NGGEMARWKCITIGETIKAGRIYMLGSTLYQPLC---TTYPDPREVTGELLETICRDEG 190
Qy 144 ---DSNQSQNRAGNSRNLGPHGLGNHHPNPNLEGRAMSALSSFQGNARTTTTST 199
Db 191 ELDIKGLMDPQWKMPRTGRTIPEIA-----PTNVEGYLLT-----226

```
QY 200 PWSRRIOHQAVIEPPMIKERRTMSXSELTEKGVNCAIPALGEVLYITRPLIYVLFIRK 259
Db 227 -----KVLRSVDVRPNLLSR-----LDNMGV-----VAELLSILRPLIYACLLFR 268
QY 260 YGVR-----WIPWALSISVDLTGMLLANSKWKGEKSKQVHFSGP--- 300
Db 269 QHVNKTVPASTKSKFPLNSPWPAPWIIIGLVIEALSRRKMGSG---WLLRQRO---SCKTPT 322
QY 301 --EKDELRRKLIALYLMRDPFTTKYTRQKLESSOKKLELPIGLFLEKIVELL 354
Db 323 ALQOMEVKGRTNLGLWMLFRGEFYQATYRPLYSIVARLEKIPGLGIFGALISDYL 378

RESULT 2
DIH9_HUMAN STANDARD; PRT; 4486 AA.
AC QN9C9; QN9C28; Q95494;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYLINDR DYNEIN HEAVY CHAIN (AXONEMAL DYNEIN HEAVY CHAIN) (DYNEIN
DE HEAVY CHAIN 9).
GN DHA9 OR DHAH17L OR DNALI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Reed W., Moats-Staats B.M., Carson J.L., Leigh M.W., Collier A.M.;
RA "A ciliary dynein heavy chain whose expression is upregulated in
RA differentiating airway epithelium.";
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE-Nasal epithelium;
RA Bartoloni L., Blouin J.L., Maiti A., Sainsbury A., Rossier C.,
RA Gehrig C., She J.X., Marron M.P., Lander E., Meeks M., Chung E.,
RA Jorissen M., Scott H.S., Delozier Blanchet C.D., Gardiner M.,
RA Antonarakis S.E.;
RA "Axonemal beta heavy chain dynein DHA9: cDNA sequence, genomic
RA structure and investigation of its role in primary ciliary
RA dyskinesia.";
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1874-1974 FROM N.A.
RC TISSUE-Nasal polyps;
RA Maiti A.K., Mattel W.G., Jorissen M., Volz A., Ziegler A.,
RA Bouvagnet P.;
RA "Chromosomal localization of human dynein heavy chain genes.";
RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=20558134; PubMed=11104725;
RA Reed W., Carson J.L., Moats-Staats B.M., Lucier T., Hu P.C.,
RA Brighton L., Gambing T.M., Huang C.H., Leigh M.W., Collier A.M.;
RA "Characterization of an axonemal dynein heavy chain expressed early in
RA airway epithelial cillogenesis.";
RA Am. J. Respir. Cell Mol. Biol. 23:734-741(2000).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF RESPIRATORY CILIA. PRODUCES
CC FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES. DYNEIN HAS ATPASE
CC ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
or send an email to license@isb-sib.ch.
CC EMBL; AF257737; AAF69004.1; -
DR EMBL; AJ404468; CAB94756.1; -
DR EMBL; AJ132088; CAA10561.1; -
DR MIM; 603338; -
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 381 410 COILED COIL (POTENTIAL).
FT DOMAIN 504 529 COILED COIL (POTENTIAL).
FT DOMAIN 539 662 COILED COIL (POTENTIAL).
FT DOMAIN 752 823 COILED COIL (POTENTIAL).
FT DOMAIN 1326 1355 COILED COIL (POTENTIAL).
FT DOMAIN 3051 3154 COILED COIL (POTENTIAL).
FT DOMAIN 3285 3341 COILED COIL (POTENTIAL).
FT DOMAIN 3640 3675 COILED COIL (POTENTIAL).
FT NP_BIND 1870 1877 ATP (POTENTIAL).
FT NP_BIND 2151 2158 ATP (POTENTIAL).
FT NP_BIND 2478 2485 ATP (POTENTIAL).
FT NP_BIND 2825 2832 ATP (POTENTIAL).
FT CONFLICT 2505 2505 L -> V (IN REF. 2).
FT CONFLICT 3678 3678 T -> A (IN REF. 2).
FT CONFLICT 4374 4374 I -> M (IN REF. 2).
SQ SEQUENCE 4486 AA; 511927 MW; 996EDFFDEB0B3EB1 CRC64;

Query Match 5.0%; Score 96.5; DB 1; Length 4486;
Best Local Similarity 21.1%; Pred. No. 17;
Matches 74; Conservative 44; Mismatches 136; Indels 97; Gaps 17;

QY 56 HIENAPTRGH-----VGSGNDPSLSYPLLIALLKDETVVEVAAEHFY----- 101
Db 2809 HINRILESPRGNALLVGVGSGKQ---SLRLAFAFISMD-VFQITLRKGYQIDRFMDL 2864
QY 102 -----GDKWNYI-ILTEAMKAVIRLALFRNSCYKMLLOGGETPNEEKDSNQSQR 153
Db 2865 ASLCIKAGVKNLNTVFLMTDAQVADERFLYLND---LLASGEIPDLYSD---DEVENI 2917
QY 154 AONSGRNLGPHGLNQHNNPNWL-----EGRAMSALSFGQNARTTSTGP----- 201
Db 2918 ISNVRNEVKSQLVD-NRENCWKFFIDRIRQLKVTLCFSPVGNKLRVRSRFFAIVNCT 2976
QY 202 -----W-----SRIOHQAVIEPPMIKERRRRTMSXSELTEKGVNCAIPALGEVLY 246
Db 2977 AIHWFEHPQQALESVSURFLQNTGIEPTKQSIKSFMAFVHTSVNQTOSYLSNEORY 3036
QY 247 -ITRPLIYVLFIRKYGVRSWI---PWAISLSVDLTGMLLANSKWKGEKSKQVHFSGPEK 302
Db 3037 NYTPKSFLEFIRLY-QSLLHRRKELKCKTERLENGLL-----KLHSTSAQV 3083
QY 303 DELRRRKLIALYLMRDPFTTKYTRQKLESSOKKLE---LIPLIGFLEKI 350
Db 3084 DDLK-----AKLAAQAEVELKQKNEADKLIQVGVGVEDKV 3118

RESULT 3
YD94_METJA STANDARD; PRT; 987 AA.
AC Q58789;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1394.
GN MJ1394.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
```


OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (FEMALE AND MALE ISOFORMS).
RC TISSUE=Pupae, and Larva;
RX MEDLINE=89168451; PubMed=2493994;
RA Burtis K.C., Baker B.S.;
RT "Drosophila doublesex gene controls somatic sexual differentiation by
RT producing alternatively spliced mRNAs encoding related sex-specific
RT polypeptides.";
RL Cell 56:997-1010(1989).
RN [2]
RP SEQUENCE FROM N.A. (MALE ISOFORM).
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K.J., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP DNA-BINDING.
RX MEDLINE=91330881; PubMed=1907913;
RA Burtis K.C., Coschigano K.T., Baker B.S., Wensink P.C.;
RT "The doublesex proteins of Drosophila melanogaster bind directly to a
RT sex-specific yolk protein gene enhancer.";
RL EMBO J. 10:2577-2582(1991).
RN [4]
RP DNA-BINDING DOMAIN, AND MUTAGENESIS.
RX MEDLINE=93178426; PubMed=8440242;
RA Erdman S.E., Burtis K.C.;
RT "The Drosophila doublesex proteins share a novel zinc finger related
RT DNA binding domain.";
RL EMBO J. 12:527-535(1993).
CC -1- FUNCTION: CONTROLS SOMATIC SEXUAL DIFFERENTIATION. BINDS DIRECTLY
CC AND SPECIFICALLY TO THE ENHANCER FBE (FAT BODY ENHANCER) OF
CC THE YOLK PROTEIN 1 AND 2 GENES (YP1 AND YP2). THIS ENHANCER IS
CC SUFFICIENT TO DIRECT THE FEMALE-SPECIFIC TRANSCRIPTION
CC CHARACTERISTIC OF THE YP GENES IN ADULT FAT BODIES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: MALE (SHOWN HERE) AND FEMALE;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- MISCELLANEOUS: EXPERIMENTALLY SHOWN TO BIND ZINC.
CC -1- SIMILARITY: CONTAINS 1 DM DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M25292; AAA17840.1; -;
DR EMBL; M25293; AAA17841.1; -;
DR EMBL; M25294; AAA17842.1; -;
DR EMBL; AE003676; AAF54168.1; -;
DR PIR; A32372; A32372;
DR PIR; B32372; B32372;
DR HSP; P04002; IATF.
DR TRANSPAC; T00955; -;
DR TRANSPAC; T00956; -;
DR FlyBase; FBgn0000504; dsx.
DR InterPro; IPR001275; -;
DR Pfam; PF00751; DM-domain; 1.
KW Sexual differentiation; Alternative splicing; DNA-binding;
KW Transcription regulation; Nuclear protein; Zinc; Metal-binding.
FT DNA_BIND 40 104 DM.
FT DOMAIN 119 224 HIS-RICH.
FT DOMAIN 267 296 SER/GLY-RICH.
FT VARSPPLIC 398 427 YSRQHNUNYDGGELRNTFRQCG (IN FEMALE ISOFORM).
FT
FT VARSPLIC 428 549 MISSING (IN FEMALE ISOFORM).
FT MUTAGEN 47 47 C->A,H: ABOLISHES DNA-BINDING.
FT MUTAGEN 50 50 H->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 59 59 H->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 68 68 C->D,Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 70 70 C->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 91 91 R->O: ABOLISHES DNA-BINDING.
SQ SEQUENCE 549 AA; 57409 MW; 3C1B92724E4CE083 CRC64;

Query Match 4.68; Score 89; DB 1; Length 549;
Best Local Similarity 23.58; Pred. No. 4.3;
Matches 51; Conservative 16; Mismatches 78; Indels 72; Gaps 7;

QY 45 AFLGIFTTINSHIENAPTRGHHVSGNDPSLSY-----GQVFLDYCKLLEKFRYPWELMPLMVILKDA 384
DB 333 ALLPISVSVNRKNGANVPL-----GQVFLDYCKLLEKFRYPWELMPLMVILKDA 384
QY 90 ETVVEVAEAEHYGDKKNYIILTKAVIRLA--LFNSGYKMLLOGGT-----PNEE 142
DB 385 DANIEASRR-----IEEARVEINRTVAQIYNYTPMALVNGAPMYLTYPSE 433
QY 143 KDSNSESQNRAGNSGRNL-----GPHGLGNQNHNPWNLEG 179
DB 434 -----QGRGAHFTLPTQICPTPEPLALSRSPSSGFSVAVNQKPSRGSNG 485
QY 180 RAMSALSSFGONARTTSTSTFGWRRIOHQOAVIEPP 216
DB 486 TVHSAASPTMTVTMTTSTSTPTLSRRQRSRATPTTP 522

RESULT 6
ID CYAA_BORPE STANDARD; PRT; 1706 AA.
AC P15318;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE BIFUNCTIONAL HEMOLYSIN-ADENYLATE CYCLASE PRECURSOR (CYCLOLYSIN) (ACT)
DE (AC-HLY) [CONTAINS: CALMODULIN-SENSITIVE ADENYLATE CYCLASE
DE (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE) (ADENYLATE CYCLASE); HEMOLYSIN].
GN CYA OR CYAA.


```
FT MUTAGEN 298 298 H->R,P,L: LOSS OF ACTIVITY.
FT MUTAGEN 301 301 E->Q,K: LOSS OF ACTIVITY.
SQ SEQUENCE 1706 AA; 177506 MW; F00744524BDD442E CRC64;

Query Match 4.6%; Score 88; DB 1; Length 1706;
Best Local Similarity 21.7%; Pred. No. 25;
Matches 70; Conservative 39; Mismatches 104; Indels 110; Gaps 17;

QY 25 LTWLLPEKFSASIGPEAVTAF-----LGITTTTNEHIEHNAPTPRGHV----- 68
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 LNKIARAGARSAGVTGRRQFRYDGMNIGVITDP-ELEVRNALNRRHAHGAQDVQVH 298
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 GSSGNDPSLSYPLLIALLKDLTVVEVAEAEHFYGDKNWYIILTEAMKAVIRLALFRNSG 128
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 GTQNNP---FP-----EAEK-----IFVSA-----TG 320

QY 129 YKMLLQGG---ETPNEEKDSNOSQNR-A-GNSGRNLGPHGLGNQNHNPWNLGRAMSA 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 ESQMLTRGQLKEYIGQRGEGYVYENRAYGVAGKSLFDGLGAA---PGVPSGR--- 372

QY 185 LSSFGQNAHTTSSSTGWSRRHQOAVIEPPMIKERRRTMSSELLTEKGVNGALFAIGEV 244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 -SKFSPDLVETVPASGLRR-----PSLGAVERQDGYSDLGVSRSFSLGEV 420

QY 245 -----LYTRPLIYVLFIRKYGVRSWIPWALSLSVDLTGMGLANSKMWGEKSKQ 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 SDMAVEAELEMTQVLA-----GARQ-----DDAEPGVSGASAHWGQRALQ 464

QY 295 VHFSGPEKDELRRKLIALYLM 317
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 465 ----GAQAVAAQR-LVHAIALM 482

RESULT 7
DYHC_ANTCR STANDARD; PRT; 4466 AA.
AC P39057;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DYNEIN BETA CHAIN, CILIARY.
OS Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Echinometridae;
OC Anthocidaris.
OX NCBI_TaxID=7629;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91326104; PubMed=1830928;
RA Ogawa K.;
RT "Four ATP-binding sites in the midregion of the beta heavy chain of
dynein".
RL Nature 352:643-645(1991).
CC -1- FUNCTION: FORCE GENERATING. PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D01021; BAA00827.1; -
DR PIR; S17231; S17231.
DR Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
KW Colled coil.
```

```
FT DOMAIN 482 502 COILED COIL (POTENTIAL).
FT DOMAIN 627 643 COILED COIL (POTENTIAL).
FT DOMAIN 734 805 COILED COIL (POTENTIAL).
FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
FT DOMAIN 1950 1978 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3030 3134 COILED COIL (POTENTIAL).
FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
FT NP_BIND 154 161 ATP (POTENTIAL).
FT NP_BIND 1852 1859 ATP (POTENTIAL).
FT NP_BIND 2133 2140 ATP (POTENTIAL).
FT NP_BIND 2460 2467 ATP (POTENTIAL).
FT NP_BIND 2805 2812 ATP (POTENTIAL).
SQ SEQUENCE 4466 AA; 511772 MW; C465CC5C6C6D325D CRC64;

Query Match 4.5%; Score 86.5; DB 1; Length 4466;
Best Local Similarity 19.7%; Pred. No. 1.2e+02;
Matches 90; Conservative 64; Mismatches 132; Indels 171; Gaps 25;

QY 1 MEAYKQWRRREYVQSFGS-----FANGL-----TWLLPEKESA 35
Db 2706 IEAFKLVF---EYAKFFEDVDDEALKAKPNIHCHFAIGDGPYMPCATW--PE---L 2757
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 36 SEIGPEAVTAFIGITTTTNEHIEHNA-----PTPRGH-----VSSGNDPSLSY 79
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2758 NKILVEALDTYNEINAVMNLVLFEDAMQHVCRINRILESPRGNALLVGVGSGKQ---SL 2814
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 PLLTAILKDLTVVEVAEAEHFYGDKNWYIILTEAMKAVIR-----L 121
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2815 ARLASYISSLE-VFOITLRKGYGIPDLKDLATCMKAGLNIGTVFLMTDAQVSDKFL 2873
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 ALFRNSGYKMLLGGETPNEEKDSNOSQNRAGNSGRNLGPHGLGNQNHNPWNL--- 177
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2874 VLIND-----LASEIDPLFAD---DEVENIIGVRNEV--KNGLODTRENCWKFFID 2923
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 ----EGRAMSALSSFGQNAHTTSTPG-----WSRRHQOAVIEPPMIKER 221
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2924 RLRLQKLVLCFSPVGTTLRVSRKRFPAVNVNCTSIDMPHEWP-----QEALVS---VSKR 2975
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 RRTMSSELLTEKGVNGALFAIGEVLYITRPLIYVLFIRKYGVRSWIPWALSLSVDLTG 281
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2976 FLDEVELLK-----GDI-----KNSTAEFMAY-VHVSVN----- 3003
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 LANSKMWGEKSKQVHFSGPEKDELRRKLIALYLMR----- 318
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3004 -ESSKQYLTNRRYNYTTP-KSFLEQIKLYESLLAMKSKELTAKMERLENGLTKLQSTAQ 3061
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 319 --DPFTTKYTRKLESSOKKLE---LIPLIGFLTEKI 350
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3062 QVDDLKAKLASOEVELAQKEDADKLIQVGVETEKV 3098

RESULT 8
CHEA_RHIME STANDARD; PRT; 758 AA.
AC Q52880;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CHEMOTAXIS PROTEIN CHEA (EC 2.7.3.-).
GN CHEA.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Ru11/001; PubMed=7623670;
RX MEDLINE=95349401;
RA Greck M., Platzer J., Sourjik V., Schmitt R.;
```

RT "Analysis of a chemotaxis operon in Rhizobium meliloti.";
RL Mol. Microbiol. 15:989-1000(1995).
CC -|- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM
CC THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEA IS
CC AUTOPHOSPHORYLATED; IT CAN TRANSFER ITS PHOSPHATE GROUP TO EITHER
CC CHEB OR CHEI (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -|- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U13166; AAA86674.1; -;
CC HSP: P04002; LWFA.
CC InterPro: IPR000410; -;
CC InterPro: IPR002545; -;
CC InterPro: IPR002570; -;
CC Pfam: PR01584; CheW; 1.
CC Pfam: PR01627; Hpt; 1.
CC DR FRAMTS; PR00344; BCTRLSENSOR.
CC Sensory transduction; Transferase; Kinase; Phosphorylation;
KW Chemotaxis.
FT MOD_RES 46 46 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 758 AA; 81110 MW; 1947225C8673831A CRC64;

Query Match 4.4%; Score 85.5; DB 1; Length 758;
Best Local Similarity 21.1%; Pred. No. 13;
Matches 63; Conservative 45; Mismatches 98; Indels 93; Gaps 14;
QY 10 RNREYVQSGFANGTLLPKEFSAEIGPEAVTAFGLI----- 49
Db 113 RSRLIKEALANG---ELPQ--AAEASAPKTPA--GVAPAAPVNVNEEGFQVAFSFD 165
QY 50 -FTTNEHIENAP---TPRGHVGSSNDPSLSYPLLIALLKLETVVEVA----- 96
Db 166 DFETGDEPTIEPTIYIVKPKSELYAKNDATL-----LURDLRLGEMSIHCDMDT 218
QY 97 -----AEHYGDKWNYIILT---EAMKAVIRLALFRNSGYKMLIQGETPNEEK 143
Db 219 LPPLDRNPPEAYF---SMKISLTKDGEEAIRSVFEAEW--DCELDVALAGGTGVMDED 274
QY 144 DSNQ-----SESQNRAGNSGRNLGPHGLGNQHNHPWNLEGRAMSALSFGONAR 193
Db 275 LPMQPPVFDLSILEDAQAPAGKEDRAAASEG-----DPRNAAVAAQAQTASNVLQMAQ 327
QY 194 TTTSTSPGNSRRIQ-----HQAQVIEPPMIK---ERRRTMSSELLTEKGVNGALFA 240
Db 328 STARVSPENARNSSQSAQAQAQAASATPTIRVDLQKRVLINLVGLVNLQAWLS 386

RESULT 9
MCM3_SCHPO
ID MCM3_SCHPO STANDARD; PRT; 879 AA.
AC P30666;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MINICHROMOSOME MAINTENANCE PROTEIN 3.
GN MCM3 OR SPC1682.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=SP011;
RX MEDLINE=98371232; PubMed=9705504;
RA Sherman D.A.; Forsburg S.L.;
RT "Schizosaccharomyces pombe Mcm3p, an essential nuclear protein,
RT associates tightly with Nda4p (Mcm3p).";
RL Nucleic Acids Res. 26:3955-3960(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 359-434 FROM N.A.
RX MEDLINE=93087163; PubMed=1454522;
RA Coxon A., Maundrell K., Kearsey S.E.;
RT "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT early step of chromosome replication.";
RL Nucleic Acids Res. 20:5571-5577(1992).
CC -|- FUNCTION: PROBABLY INVOLVED IN THE INITIATION OF DNA REPLICATION.
CC -|- SUBUNIT: ASSOCIATES STRONGLY WITH NDA4.
CC -|- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF063864; AAC32263.1; -;
CC DR EMBL: AL031525; CAA20668.1; -;
CC DR EMBL: Z15034; CAA78752.1; -;
CC DR PIR: S25528; S25528.
CC DR PIR: S26642; S26642.
CC DR InterPro: IPR001208; -;
CC DR Pfam: PF00493; MCM; 1.
CC DR PROSITE: PS00847; MCM_1; 1.
CC DR PROSITE: PS50051; MCM_2; 1.
CC DR Transcription regulation; DNA-binding; Nuclear protein;
KW DNA replication; ATP-binding.
FT DOMAIN 306 513 MCM.
FT NP_BIND 356 363 ATP (POTENTIAL).
SQ SEQUENCE 879 AA; 97481 MW; 876051CC7DE2504A CRC64;

Query Match 4.4%; Score 85.5; DB 1; Length 879;
Best Local Similarity 21.9%; Pred. No. 16;
Matches 44; Conservative 26; Mismatches 74; Indels 57; Gaps 8;
QY 87 KLETVEVAAEHYGDKWNYIILTAMKA---VIRLALF-----RNSGYKMLIQGG 136
Db 633 RTLETILRLSTAH--AKARUSSVVEVKAKEILRYALFREVVKPKRKHKKQRUEAG 690
QY 137 -----ETPNEEKDSNQ-----SESQNRAGNSGRNL----- 161
Db 691 EEFDSEDDNSDDMIESESEEMDTNMVIDSGSRRVTRSQNTSQSGSGSEIGSSIAGTA 750
QY 162 GPHGLGNQHNHPWNLEGRAMSA-----LSFGONARTTTSTTSGWSRRIRHQQAIVTEPP- 216
Db 751 GSYNVGTSTNTQLSWPSTHSTLPATSRLEASSDRNINTGTSVASEVSASVSEQSTVSLPRE 810
QY 217 -----MKERRRTMSSELLTEK 232
Db 811 KMSVFMARLASLTKSELFSEE 831

RESULT 10
AMPM_BACHD
ID AMPM_BACHD STANDARD; PRT; 248 AA.
AC Q929U4;

DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METHIONINE AMINOPEPTIDASE (EC 3.4.11.18) (MAP) (PEPTIDASE M).
GN MAP OR BH0156
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=99209008; PubMed=10192928;
RA Takaki Y., Nakasone K., Hirama C., Takami H., Inoue A., Horikoshi K.;
RT "Sequence analysis of a 32-kb region including the major ribosomal
RT protein gene clusters from alkaliphilic Bacillus sp. strain C-125";
RL Biosci. Biotechnol. Biochem. 63:452-455(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
CC PROTEINS.
CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE +
CC PEPTIDE.
CC -1- COFACTOR: COBALT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24A; ALSO KNOWN AS THE
CC MAP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB017508; BAA75293.1; -
CC EMBL; AP001507; BAB03875.1; -
CC HSP; P07906; 1C24.
CC InterPro: IPR000994; -
CC InterPro: IPR001714; -
CC InterPro: IPR002467; -
CC Pfam; PF00557; Peptidase_M24; 1.
CC PRINTS; PR00599; MAPEPTIDASE.
CC PROSITE; PS00680; MAP_1; 1.
KW Hydrolase; Aminopeptidase; Cobalt.
FT METAL 94 94 COBALT (BY SIMILARITY).
FT METAL 105 105 COBALT (BY SIMILARITY).
FT METAL 168 168 COBALT (BY SIMILARITY).
FT METAL 201 201 COBALT (BY SIMILARITY).
FT METAL 233 233 COBALT (BY SIMILARITY).
SQ SEQUENCE 248 AA; 27155 MW; C2295A5147447327 CRC64;

Query Match 4.4%; Score 85; DB 1; Length 248;
Best Local Similarity 22.0%; Pred. No. 3.2;
Matches 44; Conservative 34; Mismatches 76; Indels 46; Gaps 9;
QY 46 FLG-ITFTNEHIEENAPTRGHVSGGNDPSLPLIAIKLDLETVEVAEHE--FY 101
DB 63 FTGICASNEELVHGIP-----GN-----RVLKEGDIISVDIGAKYNGYH 103
QY 102 GDKKNVNIITLTEAMKAVIRLALFRNSGYKMLQ--GGETPNEKDSNSESQNRAGNSGR 159
DB 104 GDSAWTYAVGNISDEQDLDLTWTSTLYKGLQAKAGARLSIAIQSAEPRGYAIR 163

QY 160 NLGPHGLGNQNHHPNWNLEGRAMSALSFGQNAARTTTSTPFGWRRRIHQOQAVTEPPMIK 219
DB 164 EYVGVGVGNLHEDP-----QIPHYG-----PPGKPRKPGMVLAIEDPMVN 205
QY 220 ERR---RTMSELLTEKGVNG 236
DB 206 AGSRVYVRTLSDNWTVTVTDG 225
RESULT 11
POLG_YEYV1
ID POLG_YEYV1 STANDARD; PRT; 3411 AA.
AC P03314; O42028;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED
DE RNA POLYMERASE (EC 2.7.7.48) (NS5)].
OS Yellow fever virus (strain 17D).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85272570; PubMed=4023707;
RA Rice C.M., Lencches E.M., Eddy S.R., Shin S.J., Sheets R.L.,
RA Straus J.H.;
RT "Nucleotide sequence of yellow fever virus: Implications for
RT flavivirus gene expression and evolution";
RL Science 229:726-733(1985).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03700; CAA27332.1; -
CC PIR; A03914; GNVVY.
CC HSP; P14336; 1SVB.
CC MEROPS; S07.001; -
CC InterPro: IPR000069; -
CC InterPro: IPR000208; -
CC InterPro: IPR000336; -
CC InterPro: IPR000404; -
CC InterPro: IPR000487; -
CC InterPro: IPR000752; -
CC InterPro: IPR001122; -
CC InterPro: IPR001157; -
CC InterPro: IPR001528; -
CC InterPro: IPR001850; -
CC InterPro: IPR002535; -
CC Pfam; PF01004; Flavi_M; 1.
CC Pfam; PF00948; Flavi_NS1; 1.
CC Pfam; PF01005; Flavi_NS2A; 1.
CC Pfam; PF01002; Flavi_NS2B; 1.
CC Pfam; PF01350; Flavi_NS4A; 1.
CC Pfam; PF01349; Flavi_NS4B; 1.
CC Pfam; PF00972; Flavi_NS5; 1.
CC Pfam; PF01003; Flavi_capsid; 1.
CC Pfam; PF00869; Flavi_glycoprot; 1.
CC Pfam; PF00949; Flavi_helicase; 1.


```
RA Meijer A.H., Souer E., Verpoorte R., Hoge J.H.C.;
RT "Isolation of cytochrome P-450 cDNA clones from the higher plant
RL Catharanthus roseus by a PCR strategy.";
RL Plant Mol. Biol. 22:379-383(1993).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. MAY BE A GRANULOL-10-HYDROXYLASE.
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -!- PATHWAY: INDOLE ALKALOID SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

01-OCT-2000 (Rel. 40, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
ENAMELIN PRECURSOR.
ENAM.
Sus scrofa (Pig).
OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus..
NCBI_TaxID=9823;
[1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP TISSUE=Enamel epithelium;
RX MEDLINE=98040070; PubMed=9372788;
RA Hu C.-C., Fukae M., Uchida T., Qian Q., Zhang C.H., Ryu O.H.,
RA Tanabe T., Yamakoshi Y., Murakami C., Dohi N., Shimizu M.,
RA Simmer J.P.;
RT "Cloning and characterization of porcine enamel mRNA.";
RL J. Dent. Res. 76:1720-1729(1997).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=97350624; PubMed=9206327;
RA Fukae M., Tanabe T., Murakami D., Dohi N., Uchida T., Shimizu M.;
RT "Primary structure of porcine 89 kDa enamel.";
RL Adv. Dent. Res. 10:111-118(1996).
CC -!- FUNCTION: PEPTIDES DERIVED FROM THE PARENT ENAMELIN ARE COMPONENTS
CC OF ENAMEL, A UNIQUE AND HIGHLY MINERALIZED ECTODERMAL TISSUE
CC COVERING VERTEBRATE TEETH.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY SECRETORY-PHASE AMELOBLASTS.
CC IMPACT ENAMELIN AND LARGE-MOLECULAR-WEIGHT ENAMELINS ARE LIMITED
CC TO THE MOST SUPERFICIAL LAYER OF THE DEVELOPING ENAMEL MATRIX,
CC WHILE LOW-MOLECULAR-WEIGHT ENAMELINS ARE OBSERVED IN DEEPER
CC ENAMELIN. PREFERENTIAL LOCALIZATION AMONG THE CHRYSALITES IN ROD
CC AND INTERROD ENAMEL.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM LATE DIFFERENTIATION TO THE
CC TRANSITION STAGE.
CC -!- PTM: PARENT ENAMELIN IS PROTEOLYTICALLY CLEAVED INTO SEVERAL
CC SMALLER POLYPEPTIDES. CLEAVAGE OF N-TERMINAL REGION OF ENAMELIN
CC OCCURS SOON AFTER SECRETION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; U52196; AAD10837.1; -
GlycoSuiteDB; O97939; -
InterPro; IPR002965; -
DR PRINTS; PR01217; PRICHEXTENSN.
KW Signal; Enamel; Glycoprotein; Hydroxylation; Phosphorylation.
FT SIGNAL 1 38
FT CHAIN 39 1142 ENAMELIN.
FT CHAIN 39 ? 56 KDA ENAMELIN.
FT CHAIN 39 ? 89 KDA ENAMELIN.
FT CHAIN 39 ? 142 KDA ENAMELIN.
FT CHAIN 39 ? 155 KDA ENAMELIN.
FT CHAIN 174 276 32 KDA ENAMELIN.
FT CHAIN 515 665 25 KDA ENAMELIN.
FT CHAIN 670 ? 34 KDA ENAMELIN.
FT CHAIN ? ? 45 KDA ENAMELIN.
FT MOD_RES 53 53 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 191 191 PHOSPHORYLATION.
FT MOD_RES 216 216 PHOSPHORYLATION.
FT MOD_RES 547 547 HYDROXYLATION.
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CONFLICT 680 680 H -> D (IN REF. 2).

DR EMBL; U52196; AAD10837.1; -
DR GlycoSuiteDB; O97939; -
DR InterPro; IPR002965; -
DR PRINTS; PR01217; PRICHEXTENSN.
KW Signal; Enamel; Glycoprotein; Hydroxylation; Phosphorylation.
FT SIGNAL 1 38
FT CHAIN 39 1142 ENAMELIN.
FT CHAIN 39 ? 56 KDA ENAMELIN.
FT CHAIN 39 ? 89 KDA ENAMELIN.
FT CHAIN 39 ? 142 KDA ENAMELIN.
FT CHAIN 39 ? 155 KDA ENAMELIN.
FT CHAIN 174 276 32 KDA ENAMELIN.
FT CHAIN 515 665 25 KDA ENAMELIN.
FT CHAIN 670 ? 34 KDA ENAMELIN.
FT CHAIN ? ? 45 KDA ENAMELIN.
FT MOD_RES 53 53 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 191 191 PHOSPHORYLATION.
FT MOD_RES 216 216 PHOSPHORYLATION.
FT MOD_RES 547 547 HYDROXYLATION.
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .). (POTENTIAL).
FT CONFLICT 680 680 H -> D (IN REF. 2).

Query Match 4.3%; Score 83; DB 1; Length 524;
Best Local Similarity 18.5%; Pred. No. 13;
Matches 50; Conservative 49; Mismatches 104; Indels 68; Gaps 11;
QY 88 DLETVEV-VAEHFYGDKNYIILTEA-----MKAVIRLALFRNSGYKMLQGGTPTN 140
Db 4 DMDTIRKATAATFALVMAWVRVLDWAFTPKRIEKLRLQOQFGFNGPYRFLV--GDV-- 59
QY 141 EERDSNQSESONAGNSGRNLGPHGLGNQNHHPNWNLEGRAMSALSSFGONARTTTSSTP 200
Db 60 KESGKHQEALESKPMFENDIVPLRPHLNH-----TINTYGRNSET----- 101
QY 201 GMSRRIOHQOAVIEPPMIKERRRTMSSELLTEGVNGALFAIGVELYITRPLIYVLFIRKY 260
Db 102 -WMGRIPRIH-VMEPELIK-----EVLTHSSKYQKNFD-----VHNPVYKFL-- 142
QY 261 GYRSWIPWALSISVDTLGLMGLANSKWKGEKSKQVFSPEKDELRRRLKLIWALYLMRDP 320
Db 143 -----TGVSFEQAKW-----SKHRIISPFAFTLEKLKGLMFLPAFAICYHD 182
QY 321 FFTKYTRQKLESSOKKLEIPLIGFLTEKIV 351
Db 183 MLTKWEKIAEKQGSHEVDIPTFDVLTSVI 213
RESULT 15
ENAM_PIG
ID ENAM_PIG STANDARD; PRT; 1142 AA.
AC O97939;
DT 01-OCT-2000 (Rel. 40, Created)

FT CONFLICT 838 840 RDH -> TTI (IN REF. 2.).
SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC5FC6 CRC64;

Query Match 4.38; Score 83; DB 1; Length 1142;
Best Local Similarity 18.7%; Pred. No. 39;
Matches 34; Conservative 15; Mismatches 65; Indels 68; Gaps 5;
QY 54 NEHIIENAPTPRGHVGSSGNDPSLSYPLLIAILKDLETWVEVAAEHFYGDKKWNYIILTE 113
Db 629 NPHYPMTDPKRTIPYNEEDP-----IDPTGDEHFPGQSRWD----- 666
QY 114 AMKAVIRLALFRNSGYKMLLOGETPNEEKDSNQSESNRAGNSGRNLGPHGLGNON--- 170
Db 667 -----MEELSPK-----EDPTVRHYEGEQTSTNQPKVLPYSLDNPSKPR 706
QY 171 -----HHNPWN-----LEGRAMSALSSFCQONARTTSSTPGHSRRIQ 207
Db 707 EDPLYGEFYFWNPENFPNTAPTVSPPVESRGYANNVAGQEEESTMFPWSWDPRIQ 766
QY 208 HQ 209
Db 767 AQ 768

Search completed: July 11, 2001, 11:36:17
Job time: 76 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 00:15:10 ; Search time 2011.93 Seconds
(without alignments)
11401.327 Million cell updates/sec

Title: US-09-545-072A-1
Perfect score: 1483
Sequence: 1 attgcaccagagagagaaaa.....attaaaaaaaaaaaaaaa 1483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_in1.*
- 5: gb_in2.*
- 6: gb_in3.*
- 7: gb_om.*
- 8: gb_ov.*
- 9: gb_pat1.*
- 10: gb_pat2.*
- 11: gb_ph.*
- 12: gb_pl1.*
- 13: gb_pl2.*
- 14: gb_pl3.*
- 15: gb_pl4.*
- 16: em_ba1.*
- 17: em_ba2.*
- 18: em_fun.*
- 19: em_htgo_hum.*
- 20: em_htgo_inv.*
- 21: em_htgo_rod.*
- 22: em_htg_hum1.*
- 23: em_htg_hum2.*
- 24: em_htg_hum3.*
- 25: em_htg_hum4.*
- 26: em_htg_hum5.*
- 27: em_htg_hum6.*
- 28: em_htg_hum7.*
- 29: em_htg_hum8.*
- 30: em_htg_inv1.*
- 31: em_htg_inv2.*
- 32: em_htg_other.*
- 33: em_htg_rod.*
- 34: em_hum1.*
- 35: em_hum2.*
- 36: em_hum3.*
- 37: em_hum4.*
- 38: em_hum5.*
- 39: em_hum6.*
- 40: em_hum7.*
- 41: em_in.*
- 42: em_om.*
- 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pl.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_vil.*
- 59: gb_vil2.*
- 60: gb_htg1.*
- 61: gb_htg2.*
- 62: gb_htg3.*
- 63: gb_htg4.*
- 64: gb_htg5.*
- 65: gb_htg6.*
- 66: gb_htg7.*
- 67: gb_htg8.*
- 68: gb_htg9.*
- 69: gb_htg10.*
- 70: gb_htg11.*
- 71: gb_htg12.*
- 72: gb_htg13.*
- 73: gb_htg14.*
- 74: gb_htg15.*
- 75: gb_htg16.*
- 76: gb_htg17.*
- 77: gb_htg18.*
- 78: gb_htg19.*
- 79: gb_htg20.*
- 80: gb_htg21.*
- 81: gb_htg22.*
- 82: gb_htg23.*
- 83: gb_htg24.*
- 84: gb_htg25.*
- 85: gb_pr1.*
- 86: gb_pr2.*
- 87: gb_pr3.*
- 88: gb_pr4.*
- 89: gb_pr5.*
- 90: gb_pr6.*
- 91: gb_pr7.*
- 92: gb_pr8.*
- 93: gb_pr9.*
- 94: gb_rol.*
- 95: gb_rod.*
- 96: gb_in4.*
- 97: gb_pr10.*
- 98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1483	100.0	1483	12	AF085354 Arabidops
2	739.6	49.9	89794	12	AC004665 Arabidops
3	48.6	3.3	318221	84	AF049184 Plasmodiu
4	45.6	3.1	3791	6	D50338 Dictyosteli
5	44.2	3.0	1141	10	AX083744 Sequence
6	44.2	3.0	3747	97	U42766 Human neuro
7	44.2	3.0	140658	8	AF112374 Danio rer
8	44.2	3.0	183511	62	AC012498 Homo sapi

```

c 9 43.6 2.9 879 53 CNS01JRG AL147405 Anopheles
10 42.2 2.9 205429 60 AC005506 AC005506 Plasmid
11 42.8 2.9 759 53 CNS06QXV AL11257 T7 end of
12 42.8 2.9 2556 89 AF125948 Homo sapi
13 42.8 2.9 3308 89 AF136629 Homo sapi
14 42.8 2.9 66441 96 PFMAL1P4 AL031747 Plasmid
15 42.8 2.9 195021 93 HSA557H15 AC006281 Human DNA
16 42.4 2.9 199551 60 AC006281 AC006281 Plasmid
17 42.2 2.8 192581 84 PFMAL13P1 AL049180 Plasmid
18 42 2.8 1120 93 HSM801270 AL122110 Homo sapi
19 41.8 2.8 1137 89 AF116707 Homo sapi
20 41.8 2.8 2945 15 SDOX1IRF1 X95973 S. douglasii
21 41.8 2.8 154679 67 AC022019 Homo sapi
22 41.8 2.8 190005 88 AC068061 AC068061 Homo sapi
23 41.6 2.8 110000 84 PFMAL13P2_2 Continuation (3 of
24 41.6 2.8 139704 81 AL513523 Homo sapi
25 41.4 2.8 1173 53 CNS061LV AL404705 T7 end of
26 41.4 2.8 1478 89 AK025967 AK025967 Homo sapi
27 41.4 2.8 40600 6 CEF20D1 278542 Caenorhabdi
28 41.4 2.8 164548 83 CEY70D2 AL008880 Caenorhab
29 41.4 2.8 192029 79 AL161935 Homo sapi
30 41.4 2.8 196646 80 AL365329 Mus muscu
31 41.4 2.8 259487 80 AL365322 Mus muscu
32 41.2 2.8 11829 4 AE001376 AE001376 Plasmid
33 41.2 2.8 245802 60 AC006279 AC006279 Plasmid
34 41 2.8 705 93 HSM800237 AL049452 Homo sapi
35 41 2.8 1466 89 AF130055 AF130055 Homo sapi
36 41 2.8 2149 93 HSA276801 AJ276801 Homo sapi
37 41 2.8 2149 93 HST000030 AL162131 Homo sapi
38 41 2.8 2191 94 BC004000 BC004000 Mus muscu
39 41 2.8 67970 96 PFMAL1P3 AL031746 Plasmid
40 41 2.8 178353 62 AC012334 AC012334 Homo sapi
41 40.8 2.8 895 53 CNS06FSV AL396821 T7 end of
42 40.8 2.8 2417 93 HSM800967 AL117457 Homo sapi
43 40.8 2.8 26203 84 PFMA13P7 AL096784 Plasmid
44 40.8 2.8 83450 12 AB004054 AB009054 Arabidops
45 40.8 2.8 152409 84 PFMAL1P1 AL031744 Plasmid

```

ALIGNMENTS

```

RESULT 1
AF085354 18-MAY-1999
LOCUS Arabidopsis thaliana shrunk seed protein (SSE1) mRNA, complete
DEFINITION
ACCESSION AF085354 mRNA PLN
VERSION AF085354.1 GI:4837732
KEYWORDS cds.
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1483)
AUTHORS Lin, Y., Sun, L., Nguyen, L.V., Rachubinski, R.A. and Goodman, H.M.
TITLE The Pex16p homolog SSE1 and storage organelle formation in
Arabidopsis seeds
JOURNAL Science 284 (5412), 328-330 (1999)
MEDLINE 99212146
PUBMED 10195899
REFERENCE 2 (bases 1 to 1483)
AUTHORS Lin, Y., Sun, L., Nguyen, L.V. and Goodman, H.M.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1998) Molecular Biology, Massachusetts General
Hospital, 50 Blossom Street, Boston, MA 02114, USA
FEATURES
source
1..1483
/organism="Arabidopsis thaliana"
/cultivar="C24"
/db_xref="taxon:3702"
/chromosome="2"

```

```

/map="near ve019"
1..1483
/gene="SSE1"
122..1225
/gene="SSE1"
/function="required for seed maturation"
/codon_start=1
/protein_id="AAD30661.1"
/db_xref="GI:4837733"
/translation="MEAYKQWNRNREYVSGFANGLTWLLPEKESASEIGPEAVT
AFLEGTITNEHIAVPRGHVSGSGNDPSLYPLLIALLKDLQETVVAEAEHFYV
DKWNYILTEAMKAVILALFRNSGYKMLQGGTPEEKDSNQSSQVRAAGNSGRN
LGPGLGNQHNHPNWNLEGRAMSALSSFGONARTTSTGWSRRIOHQOAVTEPPMI
KERRTNSQLTEKGVNGALFAICEVLYITRPLIYVLFIRKYGVRSIPWALISVDT
LGMGLLANSKWKGEKSKQVHSGPEKDELRRKLIALYLMRDPFFKYTRQKLESSQ
KKLEFLPLGLTEKIVELLEGAQSRITISGS"
BASE COUNT 461 a 268 c 347 g 407 t
ORIGIN
Query Match 100.08; Score 1483; DB 12; Length 1483;
Best Local Similarity 100.08; Pred. No. 0;
Matches 1483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attgcaaccaggaagaaataacagagattgatttaacgtgaatggaattgttg 60
Db 1 ATTGCAACCAAGAGAGAAAGAAATCAGAGATTGATTACGTGAATGGAATTTGTG 60
QY 61 ttcccaattctctgsgaataagcaagtcagttgtttctctctctctctgaagtc 120
Db 61 TTTCCCAATTTCTCTGAGAAATAGCAAAAGTTCAGTTTGTGTTCTCTATCTGAAGTC 120
QY 121 aatggaactataagcaatgggttgagaaatagagagatgatacaatccttgagtc 180
Db 121 ATGGAACCTTATAGCAATGGGTTTGAGAAATAGAGAGATGTACAAATCCTTTGGATC 180
QY 181 cttgtccaacgattgacatggctcttcctcctgagaagtttctctcctcagagattggacc 240
Db 181 CTTTCCCAACGATTGACATGGCTGCTTCTCTGAGAAGTTTCTGCTTCAGAGATTGGACC 240
QY 241 agaagcagtaacggcttttttgggcataatcacaacgataataagaacacataatgaaaa 300
Db 241 AGAAGCAGTAACGGCTTTTGGGCATATTCACAACGATAAATGAACACATAATTTGAAA 300
QY 301 tgcctcaacactcgtgcccattgttgatcttcggaagatgataccatccctcttctatcc 360
Db 301 TGCTCCACACACTGCTGGCCATGTTGGATCTTCCGGGAATGATCCATCCCTTTCTTATCC 360
QY 361 actactcactcctcctcaaggatttggaactgttggaagtgtggaagtggcagctgaacactt 420
Db 361 ACTACTCATCGCCATCTCTCAAGGATTGGAACATGTTGTGGAAGTGGCAGCTGAACACTT 420
QY 421 ctatggagacaaaaatggaactacattattctcactgaagctatgaagcgtctcatag 480
Db 421 CTATGGAGACAAAAATGGAACATACATTATTCTCACTGAAGCTATGAAGGCTGTCATTAG 480
QY 481 gtagcctgttcggaatagtggtgataagatgcttctcaaggagggaacacacataa 540
Db 481 GTTAGCCTTGTTCGGGAATAGTGGGTATAGATGCTTCTTCAAGGAGGGGAACACCTAA 540
QY 541 tgaggagaagattctaaccaatccgagtcgcaaaatagagctggtgtaattcgggtagaaa 600
Db 541 TGAGGAGAAAGATTCTAACCAATCCGAGTCGCAAAATAGAGCTGTAATTCGGGTAGAAA 600
QY 601 tctcgggcctcattgcttggaaacacaaatcatcataatccatggaacttgggaagacg 660
Db 601 TCTCGGCCCTCATGGTCTTGGAACCAAAATCATCATATCCATGGAACCTGGAAGACG 660
QY 661 ggcgatgtctgcttaagtctatttggtcagaaatgcaagaacacacacacatctctacccc 720
Db 661 GGCATGTCTTAAAGTTCAATTTGGTCAGATGCAAGATGCAAGAACACACATCTTCTACCCC 720

```

```
Qy 721 cggttggtcgaagaattcaacatcagcagcttagatagagctccaatgatcaagga 780
|||||
Db 721 CGTTGGTCTCGAAGAAATCAACATCAGCAGCAGTATAGAGCTCCATGATCAAGGA 780
Qy 781 gagcggaagaacgatgtccgagctactactagaaaggtgttaatgagcgtgtttgc 840
|||||
Db 781 GAGCGGAAGAACGATGTCGAGCTACTTACTTGAGAAGGCTTAATGAGCGTGTGTTGC 840
Qy 841 gattggtgagtttttatacaagaagacgcctcatcttaacgtttcttcatcagaata 900
|||||
Db 841 GATTGGTGAGGTCTTTACATAAGAGACCGCTCATTTACGTCTCTTTTCATCAGAAAAA 900
Qy 901 tggagtcgagcttggtatctctggcttatcgtcttcttgagacacactgggagtg 960
|||||
Db 901 TGGAGTCGAGCTTGGATTCCTTGGCTATATCGCTTCTCTGGACACACTGGGATGG 960
Qy 961 tcttcttgaattgaagtggtgggagagaagcaagtcacattcttctcaggacc 1020
|||||
Db 961 TCCTTCTTCAAAATCGAAGTGTGGGAGAGAGCAAGCAAGTCCATTTCTCAGGACC 1020
Qy 1021 tgaagaagatgagctgagagacacaaaactatgagcattgtacctcatgagagatcc 1080
|||||
Db 1021 TGAAGAAGATGAGCTGAGAGACGAAACTGATATGGGCTATGTACCTCATGAGAGATCC 1080
Qy 1081 attcttcaccaagtacacaaagcagaagctggaagctctcaaaagaagctggaactaat 1140
|||||
Db 1081 ATTCTTCAACCAAGTACACAAAGCAGCAAGCTGGAAGCTCTCAAAAGAGCTGGAACATA 1140
Qy 1141 tccattgagctgattctcctacagagaagattgtggagcttttgaggagctcagtcacg 1200
|||||
Db 1141 TCCATTGATGCTGATTCCTCACAGAGAAGATTGTGGAGCTTTTGGAGGAGCTCAGTCAGC 1200
Qy 1201 gtacacttacatctggatcgtgaagctgaagcttttacttatgttttatgatacagc 1260
|||||
Db 1201 GTACACTTACATATCGGATCGTGAAGTAAAGCGTTTACTTATGTTATGTTATGCAACGG 1260
Qy 1261 aagaattatgcattgttggagctgttttagatcatcaagagctcctacagattctt 1320
|||||
Db 1261 AAGATATTTGCCATTGTGGAAATGCTTTTATGATCATCAAGGCTCCTCAGATTCTT 1320
Qy 1321 aggaatagtttcagctttgttagaattgtgtttatttcaacagctagagaacataa 1380
|||||
Db 1321 AGGGAATGGTTTCAGGCTTTGTAGAAATTTGTGTTATTCACACAGGTAGAGAACATAA 1380
Qy 1381 ccatagacagatgtatctgaagagataagctctctatgtctaaagaataagaccgatac 1440
|||||
Db 1381 CCATAGACAGATGATCTGAGAGATAGCTTCTCTATGTCTAAAGAATGGACCCATAC 1440
Qy 1441 gaataaaacagcatcattaaagattaaaaaataaaaaa 1483
|||||
Db 1441 GAATAAAACAGCATCAATTAAGATTAAAAAATAAAAAAAAAA 1483
```

```
RESULT 2
AC004665/c 89794 bp DNA PLN 05-APR-2000
LOCUS Arabidopsis thaliana chromosome II section 244 of 255 of the
DEFINITION complete sequence. Sequence from clones F17K2, F4118.
ACCESSION AC004665 AE002093
VERSION AC004665.2 GI:6598428
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 89794)
Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D.,
Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E.,
Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M.,
Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L.,
Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H.,
Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,
```

Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.
Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
Nature 402 (6763), 761-768 (1999)
20083487
10617197
2 (bases 1 to 89794)
Lin, X.
Direct Submission
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Dec 17, 1999 this sequence version replaced gi:3386593.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL
(ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
University of Washington), Genscan (Chris Burge,
http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/services/NetGene2/), searches of the
complete sequence against a peptide database and plant EST
databases at TIGR, and manual curations based on those analyses.
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by two
or more gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted as tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
F6223, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
and Satoshi Tabata for helpful assistance. In addition, we would
like to thank the TIGR Bioinformatics Department, especially Lixin
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
Peterson, Michael Holmes, and Delwood Richardson for software and
database support.

This work was supported by the National Science Foundation,
Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES	Location/Qualifiers
source	1..89794 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="II"
misc_feature	<1..2743 /note="Sequence from clone F17K2" complement(join(<940..1001,1102..1163,1345..1412, 1457..2051,2136..2355,2646..>2721)) /gene="At2g45690" complement(<940..>2721) /gene="At2g45690"
mRNA	/note="F17K2.22; predicted by genscan and genefinder" complement(join(940..1001,1102..1163,1345..1412, 1457..2051,2136..2355,2646..>2721)) /gene="At2g45690"
gene	/note="hypothetical protein" /codon_start=1 /protein_id="AAC28524.2" /db_xref="GI:6598429"
CDS	

```
/translation="MFVLIQGLTWLIPKFSASETQPEAVTAPLGIPTSTINHHIENA
PTRRGVSGNDPSLSYPLLTAILDKLETVEVEAAEHFYGDKNWYIILTAMKAVI
RLALFRNSGYKMLQGETPNEEKDSNQESNRAGNSRLNPGHLGNQNHNPWL
EGRAMSALFGONARTTSSTPGWRRIHQHOAVTEPPMKERRTMSSELLTEKGVN
GALFATGEVLYITRPLIYVLFIRKYGVRSWIPWALSISVDITLGMGLLANSKWKGEKSK
QVHFSPEKDEVRNLDLNTLRRRLKLIWALYMRDPFTTKTROKLESSQKLELIP
LIGLFEKIVLEGAOSRYTISGS"
complement(2744..>89794)
/notes="Sequence from clone F4118"
2858..2975
/rpt_family="(CAAA)n"
complement(join<4691..4835,4915..5018,5130..5341,
5421..5526,5765..5872,5973..6190,6289..6352,6549..6935,
7016..7210,7348..>7980))
/gene="At2g45700"
complement(<4691..>7980)
/gene="At2g45700"
/notes="F4118.32; predicted by genscan"
complement(join(4691..4835,4915..5018,5130..5341,
5421..5526,5765..5872,5973..6190,6289..6352,6549..6935,
7016..7210,7348..7980))
/gene="At2g45700"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAC28555.1"
/db_xref="GI:3386625"
/translation="MSNTVEDDDDDFOIPPSSQLSIRKLPHPTNANNTSHRPPNKKPR
LCRYPCKEVTPPPDPDLFCSSSTPHCLIDCIPSSVDCSLGDFNGPSTSSIGEDKE
DKDCCIKVNRGVLCSMEARLLKSRICIGFDPSGLHEDDGEVSNSELDVLINLCS
SEGRSEFSLGRKDDSTQCPLCSMDISLSEERQVHNSWNTCLKSNQPSQESLKRCE
NLSLIGESIDDDVQPLQVLTDLSPVLKWLRLSLGAKYEDVFIREIDMDTILQSLTEE
DLSIGITSLGRPKRIYNALSGVRDFFASAEVQASOCHTSGHVTERQDKSTTRKAS
EPKPTANKLITTEFFQATTEGKIRTPAPKVAEKSPSDSSRRRAVRNRNNGKSKVI
PHNCCIPGPFVDAFKYLTROCCWHLFPHLDHYOGLTKFSHGKIYCSLVTAFLV
NMKIGIPWRLOVLDLQGVNLSGIDVTCFDNHCPSIMILFEPAKGLVHTGDFR
YSEMSNWLIGSHLSLIDTYICNPDPFKQEAIVQVVEAIGAEAFNPKTLFLIG
SYTIGERFLFLEVRLEKRIYINPAKLLECLGFSKDDIQMFTVKEESHVHVPL
WTLASPKRLKHWANRYNLSYLVAFSPGTSGTKKSPGRKLOQGTIIRYEPYS
EHSFTLEKFEVQKVSPEVIIPSVNNDGPDSDAAAMVSLVT"
join(8646..8720,8995..9096,9231..9259,9354..9685)
/gene="At2g45710"
8646..9685
/gene="At2g45710"
/notes="F4118.31"
join(8718..8720,8995..9096,9231..9259,9354..9474)
/gene="At2g45710"
/codon_start=1
/product="40S ribosomal protein S27"
/protein_id="AAC28554.1"
/db_xref="GI:3386624"
/translation="MYLQNDIDLNPAPAELEKRRKHLKRLVQSPNSFFMDVKQCGCFN
ITTVFHSQTVMVCGNCOTLLCTPTGGKAKLTGGCSFRKK"
<11943..>13604
/gene="At2g45720"
<11943..>13604
/gene="At2g45720"
/notes="F4118.30"
11943..13604
/gene="At2g45720"
/notes="unknown protein"
/codon_start=1
/protein_id="AAC28553.1"
/db_xref="GI:3386623"
/translation="MYEERTGNVTLIDQTVEDLLLAQELVPIALS KARTVKGFSSRW
RVISLRLEIPTCLSSHPCFSKHTLCQEOLQAVLETKETIELANVCSEKEGK
LKMOSDLSLAKIDLSLKDCLGLMKTGVLTPTLSSSTODLTFVSVELLARLOI
GHLESKRKALEQLVPMKEDERAVITAGRTNVASLVOLLTATSPSVRENATVTCSL
AESGGCNLISENLAVPLIRLESISIVAKERAVISLORMSISSETSRISVGHGGV
PLEICTGDSVSASACTLKNISAVPERQNLABEGIVKVMINILNCILGSGVEY
AAECLQNLTSSNETLRVSIENGITLQYLDGLPQESGVAAIRNLVGSVSVETFY
KIIPSLVHLKSGISGAQQAOSTICRIATNETKRMIGESCIPLLIRMLEKASGA
REVAQAQIASLVTPNCRVREKRSVTSLVMLLEPSPGNSAKKYAVYSGLAALCSSR
KCKKLNVSHGAVGLYLLKLELSEVPGSKLLERIEKGLKLSFPRK"
```

Db	799	AGGTAGAACAATCAACCATAGACAGATGATCTGAAGAGATAAGCTTCTCTATGTCTAAA	740
Qy	1426	gaatgagccatacagaataaaacaagcatcaattaaagaattaaa	1469
Db	739	GAAATGGACCGATACGAATAAACAAGCATCATTAAGATTAAA	696
RESULT	3		
PFMAL13P3/C			
LOCUS	PFMAL13P3	318221 bp	DNA HTG 19-AUG-1999
DEFINITION	Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.		
ACCESSION	AL049184		
VERSION	AL049184.5 GI:5763803		
KEYWORDS	HTG; HTGS, PHASE1.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M., and Barrell,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK		
COMMENT	On Aug 24, 1999 this sequence version replaced gi:5731882. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. * NOTE: This is a 'working draft' sequence. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.		
FEATURES	Location/Qualifiers		
source	1..318221 /organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /chromosome="13"		
BASE COUNT	113610 a 29264 c 27321 g 118405 t 29621 others		
ORIGIN			
Query Match	3.3%; Score 48.6; DB 84; Length 318221;		
Best Local Similarity	64.9%; Pred. No. 0.14;		
Matches	72; Conservative 0; Mismatches 39; Indels 0; Gaps 0		
Qy	1373	gaacataaccatagacagatgtatctggaagagataagttctctatgtcttaagaagaatgg	1432
Db	12715	GAATAAAAAAATATATATATATTTATATATTTATGCATTATATTTTTATTATTAAGA	12656
Qy	1433	accgatcgataaaacaagcatcattaagaattaaaaaiaaaaaaaaaa	1493
Db	12655	AACGACACAAAAAIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	12605
RESULT	4		
D50338/c			
LOCUS	D50338	3791 bp	DNA INV 14-APR-2000
DEFINITION	Dictyostelium discoideum DNA for ribosomal protein, complete cds.		
ACCESSION	D50338		
VERSION	D50338.1 GI:790953		
KEYWORDS	ribosomal protein.		
SOURCE	Dictyostelium discoideum (strain:AX-3) vegetative growth amoeba cell DNA, clone.lib:HincII limited lambda gt10 clone:lambda GH1.		
ORGANISM	Dictyostelium discoideum		

REFERENCE 1 (bases 1 to 3791)
 AUTHORS Ohmachi, T., Kimura, Y., Asada, Y., and Ennis, H. L.
 TITLE A Dictyostelium discoideum developmentally regulated gene encoding a ribosomal protein homologous to rat L27a
 JOURNAL Unpublished (1995)
 REFERENCE 2 (bases 1 to 3791)
 AUTHORS Ohmachi, T.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-1995) to the DDBJ/EMBL/GenBank databases. Tetsuo Ohmachi, Hiroaki University, Science of Bioresources; 3 Bunkyo-cho, Hirotsaki, Amori 036, Japan (Tel:0172-36-2111(ex.4634), Fax:0172-35-9000)

FEATURES
 source 1..3791
 /organism="Dictyostelium discoideum"
 /strain="AX-3"
 /db_xref="taxon:44689"
 /cell_type="amoeba cell"
 /clone="lambda GH1"
 /clone_lib="HincII limited lambda gt10"
 /dev_stage="vegetative growth"
 /dev_stage="vegetative growth"
 2327..2349
 /note="putative"
 /number=1
 join(2347..2349,2661..3104)
 /codon_start=1
 /product="ribosomal protein"
 /protein_id="BAA08873.1"
 /db_xref="GI:790954"
 /translation="MPTFRSKRHSRGDVCAGYGRVGRKHKRHPGGRGNAGGLTHRRIN
 FQYHPGFGKLMRHFHLRNQYHCPTVSLK IWTLPESVRSKSLAAKNDGTPVVD
 VTQKGFYVLGHILPTQPIIVKARYFSKVAEKKIKAVGGACILVA"
 intron 2350..2660
 exon 2661..3168
 /number=2
 3'UTR 3105..3168
 polyA_signal 3121..3126
 polyA_signal 3131..3136
 BASE COUNT 1492 a 464 c 459 g 1376 t
 ORIGIN

Query Match 3.1%; Score 45.6; DB 6; Length 3791;
 Best Local Similarity 57.9%; Pred. No. 0.42;
 Matches 81; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
 QY 1344 tagaaattgtttattgcaacaggtagagaaacataaccatagacagatgtatctgaaga 1403
 Db 1885 TAATAATATTTCATATTAATAATTTTGAATAAATTTGTATATAGATAAATAAGTAAT 1826
 QY 1404 gataagctctctatgctctaaagaatggaccgatcacgaataaacaagcatcattaaag 1463
 Db 1825 AATAAGTTTATTTCTATTTTAAATTTGTAGACTAACACCAAAAAAAAAAAAAAAAAA 1766
 QY 1464 attaaaaaataaaaaa 1483
 Db 1765 AAAAAAAAAAAAAAAAAA 1746

RESULT 5
 AX083744/c 1141 bp DNA PAT 28-FEB-2001
 LOCUS AX083744
 DEFINITION Sequence 22 from Patent WO0111061.
 ACCESSION AX083744
 VERSION AX083744.1 GI:13185472
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM Dictyostelium discoideum
 REFERENCE 1 (bases 1 to 1141)
 AUTHORS Kunst, L. and Clemens, S.

TITLE Regulation of embryonic transcription in plants
 JOURNAL Patent: WO 011061-A 22 15-FEB-2001;
 UNIVERSITY OF BRITISH COLUMBIA (CA)
 FEATURES
 source Location/Qualifiers
 1..1141
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 promoter 1..1141
 /note="consensus sequence of A. t., L. a., and B. n. FAEL
 promoters"
 BASE COUNT 123 a 32 c 42 g 112 t 832 others
 ORIGIN

Query Match 3.0%; Score 44.2; DB 10; Length 1141;
 Best Local Similarity 10.8%; Pred. No. 0.81;
 Matches 86; Conservative 277; Mismatches 432; Indels 1; Gaps 1;
 QY 3 tgaacaccaggaagaagaataacagagattgatttaacgtgaatggaattgtgtt 62
 Db 1006 KSCATNNAMWYATTWAAAYAAAKWARWAGNNRMVYGAAGNKGWCMAMATMBGWADTA 947
 QY 63 tcccaaatctctgagaataagcaagttcagttttctctctctctctctgaagctcaa 122
 Db 946 GKMNNNNNNNTDVRWAMAKAKNNNNNNNAYWYACYNRAAATNNKMMHMKWTHGASKR 887
 QY 123 tggagcctataagcaatgggtttgagagaatagagatgatgacaatcccttggatcct 182
 Db 886 RTRHRTCR-RTKYNNNNNNNARTVYVYHAAARRMMNAWTRTNNNNNNNNNACRNTRT 828
 QY 183 ttgcaacaggattgacatggctctctctgagagatttctgctcagagattggaccag 242
 Db 827 WWABWKHSWCNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAARMAARTCNVYHAA 768
 QY 243 aagcagtaacgctttttggccatattcacaacagataaataagacaataattgaaagt 302
 Db 767 VTTTHTDMCYTMMWTYDMMYMBTTTTTNNMTTSTNTNNNNNNNNNNNNNNNNNNNN 708
 QY 303 ctcccaacacacctctggccatgttgatctctcgggaatgatccatccctctcttaccac 362
 Db 707 AYYAHATNNWGCWNNNTDARRTNTTVMRRWMTNTKTRWYSTTRRHHTGATNNNNNN 648
 QY 363 tactcatgcctcctcagagatttgaaactgtgtgagaagtgagcagctgaacactctc 422
 Db 647 NNNNNNNNSCCTCTRMNTMTMKDGMTVRKKYKWRDITCTTYVDVWADSWVWYANMM 588
 QY 423 atggagacaaaaatggacatattctcactgaagctatgaagcgtgtcattaggt 482
 Db 587 RCRDYTYRNTYCKSYAHSYVWNSNNAMWYRRYSARWSSMARWTTNNNNNNNNNNNN 528
 QY 483 tagcctgttccggaatagtggtgataagatgcttctcctcagaggagggaacacactaatg 542
 Db 527 WAGTMMWRHNNNTDTRY YWKKRWARBTTVYDSMCNAKSMWRGNWRAKMWAAANN 468
 QY 543 aggaagaagattctaaacacacccagctgcgaataatagagctggtaattcgggtgaaatc 602
 Db 467 DAGAMDHTYMGNTTMMWRRAKMMNNMNAWCCRAYCCNNNNNNNNNNNNNNNNNNNN 408
 QY 603 tcgggctcctgctcttggaacacacacacacacacacacacacacacacacacacac 662
 Db 407 WKAACNN 348
 QY 663 cgatgctgtcgttaagttcatttggtcagaatgcaagaacacacacacacacacacac 722
 Db 347 WRBCKTTTSMWMMMDHMTHTCTYGNNTWGSAYBMAAASBMAAASBMAAASBMAAAS 288
 QY 723 gttggtctcgaagaattcacaacacagcaggttatagacccctcccaatgatcagaaga 782
 Db 287 TMTNN 228
 QY 783 ggcgaagaacgatgtc 798
 Db 227 WRRABHRSNN 212


```
RESULT 6
HSU42766 3747 bp mRNA PRI 22-JUL-1996
LOCUS Human neurotrophin y2 receptor mRNA, complete cds.
DEFINITION
ACCESSION U42766
VERSION U42766.1 GI:1174045
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Gehlert,D.R., Beavers,L.S., Johnson,D., Gackenheimer,S.L.,
Schobert,D.A., and Gadsdill,R.A.
TITLE Expression cloning of a human brain neurotrophin Y Y2 receptor
JOURNAL Mol. Pharmacol. 49 (2), 224-228 (1996)
MEDLINE 96226058
REFERENCE
AUTHORS Gehlert,D.R.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1995) Don R. Gehlert, CNS/GU Research, Eli Lilly,
Inc., Lilly Corp. Center, Indianapolis., IN 46285, USA
FEATURES
source
1. 3747
/organism="Homo sapiens"
/db_xref="taxon:9606"
496..1641
/notes="hNPY2 receptor"
/codon_start=1
/product="neurotrophin y2 receptor"
/protein_id="AA04120.1"
/db_xref="GI:1174045"
/translation="MGPIGAEDENQTEEMKVEQCPQTPRGELVPDPPELIDST
KLEQVWVILAYCSILLGVIGNSLVIVHVKFSMTVTNFFIANLAVALLVNTL
CLPTFTVTLGMKMPVLCVLPVAVGLAVOVSTITLVIALDRHRCIVVHLESKI
SKRISLLIYVPLGLIISFYSYTRINKLNHNSPGAANDHYHORRKTVMKLVVVV
YLSLLIYVPLGLIISFYSYTRINKLNHNSPGAANDHYHORRKTVMKLVVVV
FWSVPLHAFQALDIDSDVLDLKEYLIFTVTHIAMCSFANPLLYGWMNSYRK
AFUSAFRCQRDLAIHSEVSVTPEAKKLEVRKNSGPNDSFTEATNV"
BASE COUNT 1151 a 759 c 753 g 1084 t
ORIGIN

Query Match 3.08; Score 44.2; DB 97; Length 3747;
Best Local Similarity 48.2%; Pred. No. 0.97;
Matches 124; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 1227 gttgaagcgtttactgtattgttatatgaacggaagaataattgcccattgttggaatgct 1286
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3490 GTTACATGTTGTATTAGTAATAATTAGATCCCAATTAATGATTCATTAATACATATATCT 3549

QY 1287 ttttagatcatcaaggctcctcacagattctttagggaatgggttcaggcgtttgttag 1346
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3550 TATCCATTCATTATGTCATTAATTAATAAATACCTTTTATGTAGAGGCTTTATGTTG 3609

QY 1347 aaattgtgtttatgcaacaggtagagaacataaccatagacagatgtatctgaagagat 1406
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3610 CAATTAATAAAGTGGGAATAATGAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3669

QY 1407 aagcttctctatgctaaagaataaggccgacgatacgaataaacaagcatcattaaagatt 1466
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3670 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3729

QY 1467 aaaaaaataaataaataa 1483
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3730 AAAAAAATAAATAAATAA 3746

RESULT 7
AF112374/c
LOCUS AFL12374 140658 bp DNA VRT 26-JAN-2001

Danio rerio odorant receptor gene cluster.
ACCESSION AFL12374.1 GI:6739470
VERSION
KEYWORDS zebrafish.
SOURCE Danio rerio
ORGANISM Danio rerio
REFERENCE
AUTHORS Actinopterygii; Neopterygii; Teleostei; Euteleostei; Osteichthyes;
Cypriniformes; Cyprinidae; Rasbora; Danio.
TITLE Analysis and Characterization of an Odorant Receptor Gene Cluster
in the Zebrafish Genome
JOURNAL Genomics 71 (1), 53-65 (2001)
MEDLINE 21100876
REFERENCE
AUTHORS Dugas,J.C. and Ngai,J.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Molecular and Cell Biology, University of
California at Berkeley, 265 LSA, Berkeley, CA 94720, USA
FEATURES
source
1. 140658
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
complement(<5892..>6845)
/gene="OR 5.2"
/product="odorant receptor 5.2"
complement(<5892..>6845)
/gene="OR 5.2"
complement(5892..6845)
/gene="OR 5.2"
/codon_start=1
/product="odorant receptor 5.2"
/protein_id="AAF27258.1"
/db_xref="GI:6739471"
/translation="MTFNITVTFEYLLGPELHPDYGVAGTFLFVYLTLLGGNI
FIATAYEKSLOKPTVIFCNLAVCDLTGTGVLPQAIAMFINITISFNGCFVOM
FVHYGVSLSLLIILIDIRVIAHNPFCVLTINKTIFAGGVITWATPLMALI
VYAYDEPYCATNVQIFCERNVVKLSRDVQKLLALIVVLPGLFVLS
FIATFVFNISDNQRTYKSTCTPQITICLYFVFKFLYVYDITVALSPSRIAL
SLGCFVPPVNPVMIYCFKTKIRHASVQVNL"
complement(8178..9140)
/gene="OR 5.5p"
/pseudo
complement(<11448..>12434)
/gene="OR 5.4"
/product="odorant receptor 5.4"
complement(<11448..>12434)
/gene="OR 5.4"
complement(11448..12434)
/gene="OR 5.4"
/codon_start=1
/product="odorant receptor 5.4"
/protein_id="AAF27259.1"
/db_xref="GI:6739472"
/translation="MNSVTPRVMTFNITVTFEYLLGPELHPDYGVAGTFLFV
YLLTLLGGNIIFIAIYAEKSLQKPTVIFCNLAVCDLTGTGVLPQAIAMFINIT
SNFLICVFEFLICLSNVSLIILIDIRVIAHNPFCVLTINKTIFAGGVITW
FIATFVFNISDNQRTYKSTCTPQITICLYFVFKFLYVYDITVALSPSRIAL
GPFIILCSGIFGTSVFNISDSQAKYKTFSTCPQILFICLHFLSKLVVYDVT
MPSPFRIALSLCSCLLPVPVNPVMIYCFKTKIRHASVQVNL"
complement(14265..14390)
/repeat_region
14265..14390
/rpt_type=tandem
/rpt_unit=tg
complement(<16811..>17794)
/gene="OR 5.3"
/product="odorant receptor 5.3"
complement(<16811..>17794)
/gene="OR 5.3"
complement(16811..17794)
/gene="OR 5.3"
/codon_start=1
```

```

/product="odorant receptor 5.3"
/protein_id="AAF27260.1"
/db_xref="GI:6739473"
/translation="MGSRNVEYQNTTIITEFVLVRLGPPGLPEVHGVSGALLFYLIL
VLATNIFIVMEIKSIQKPTIIFCNLAMADLALGTTTTPRVIAKYWMSDKLIISF
SCQTFQMFVFLGATNLFALMALDRFAICNPLRYHTLTKNSTLILICAFWTAN
MLQVGVLLISVYKCSNVIPTCHYCDHVSISKLACDAATMKVTSTAIAMFWLWGP
LSFISVSYIISVSKISSEGRYKFTSTCTPOLLIIICLYLPRTFVIYIINISGYEL
STDTRVVMYSLIPAVINPFIYCFYKTEIKAEIKFRKKNKVDKRISLG"
complement(<20838..>21887)
/gene="OR 5.1"
/product="odorant receptor 5.1"
complement(<20838..>21887)
/gene="OR 5.1"
complement(20838..21887)
/gene="OR 5.1"
/codon_start=1
/product="odorant receptor 5.1"
/protein_id="AAF27261.1"
/db_xref="GI:6739474"
/translation="MYCTLKSACVNVNNAISILVASFEKLGSSVMMAYTNITE
FYLLGPGHPEYGAAGFLLFVLTLAGNIFIIISFYAEKSLQKPTLYLFCNLAV
CDIATGVTLPKAIKAYFEDSSEITFLSCFQMFVHYLGINSFLMLLMAVDRFVAV
CNPLRYCVLTNKTLLACALIMWLCISMLTWMYNTFDEPYCGSNITONCYCHNSI
IKLSGDIQKQKFSFAMFVLGGLPIIFESFALIIISVFLSDTQAEKFTSTCT
POLLIIICLYYPRCVYVVDHTIATISPGIRVWLINWLTILPIIPVPMIYCFYREIKD
AIKKLKKDKVNVQI"
complement(31126..32258)
/gene="OR13.5p"
/translation="similar to OR 13"
/pseudo
complement(<34090..>35034)
/gene="OR 13.2"
/product="odorant receptor 13.2"
complement(<34090..>35034)
/gene="OR 13.2"
complement(34090..35034)
/gene="OR 13.2"
/codon_start=1
/product="odorant receptor 13.2"
/protein_id="AAF27262.1"
/db_xref="GI:6739475"
/translation="MSVGNSTFAKDFATVGPGLQPHYGLVLSALLEFVYVLTILGNA
VFFTLFVTKSLHPRVYFIINLVASOVLFSTTLPKIIISRYFQDGTISFLGCFVOM
FVHFLGACSVLWMAIDRYTAICYPLQVHSIMNKNVLIILISPWILSFWGFLAL
VIRAPLPYCADNSIIHCYCDHVSITTLACTDRSLYSIPALIYAFLLILVPPFAVIFS
YCSIFIAVMRISGSGRMKFTFSCPOLIIIALIYFIPRVNLYSSNIGLKAFSTDLRII
MIMYSVLPPLNPIIYCLRTEEVKVLRLQHFSEKTR"
complement(<39727..>40668)
/gene="OR 13.3"
/product="odorant receptor 13.3"
complement(<39727..>40668)
/gene="OR 13.3"
/codon_start=1
/product="odorant receptor 13.3"
/protein_id="AAF27263.1"
/db_xref="GI:6739476"
/translation="MPVENSFVKDFVIVGPGIOPHYGLVLSALLFFVYVLTILGNA
VFFTLFVTKSLHPRVYFIINLVASOVLFSTTLPKIIISRYFQDGTISFLGCFVOM
FVHFLGACSVLWMAIDRYTAICYPLQVHSIMNKNVLIILISPWILSFWGFLAL
VIRAPLPYCADNSIIHCYCDHVSITTLACTDRSLYSIPALIYAFLLILVPPFAVIFS

```

```

repeat_region
/rpt_type=tandem
/rpt_unit=ca
complement(27763..28719)
/gene="OR 13.4p"
/translation="similar to OR 13"
/pseudo
complement(31126..32258)
/gene="OR13.5p"
/translation="similar to OR 13"
/pseudo
complement(<34090..>35034)
/gene="OR 13.2"
/product="odorant receptor 13.2"
complement(<34090..>35034)
/gene="OR 13.2"
complement(34090..35034)
/gene="OR 13.2"
/codon_start=1
/product="odorant receptor 13.2"
/protein_id="AAF27262.1"
/db_xref="GI:6739475"
/translation="MSVGNSTFAKDFATVGPGLQPHYGLVLSALLEFVYVLTILGNA
VFFTLFVTKSLHPRVYFIINLVASOVLFSTTLPKIIISRYFQDGTISFLGCFVOM
FVHFLGACSVLWMAIDRYTAICYPLQVHSIMNKNVLIILISPWILSFWGFLAL
VIRAPLPYCADNSIIHCYCDHVSITTLACTDRSLYSIPALIYAFLLILVPPFAVIFS
YCSIFIAVMRISGSGRMKFTFSCPOLIIIALIYFIPRVNLYSSNIGLKAFSTDLRII
MIMYSVLPPLNPIIYCLRTEEVKVLRLQHFSEKTR"
complement(<39727..>40668)
/gene="OR 13.3"
/product="odorant receptor 13.3"
complement(<39727..>40668)
/gene="OR 13.3"
/codon_start=1
/product="odorant receptor 13.3"
/protein_id="AAF27263.1"
/db_xref="GI:6739476"
/translation="MPVENSFVKDFVIVGPGIOPHYGLVLSALLFFVYVLTILGNA
VFFTLFVTKSLHPRVYFIINLVASOVLFSTTLPKIIISRYFQDGTISFLGCFVOM
FVHFLGACSVLWMAIDRYTAICYPLQVHSIMNKNVLIILISPWILSFWGFLAL
VIRAPLPYCADNSIIHCYCDHVSITTLACTDRSLYSIPALIYAFLLILVPPFAVIFS

```

```

mRNA
/gene="OR 13.1"
/product="odorant receptor 13.1"
complement(<42101..>43087)
/gene="OR 13.1"
complement(42101..43087)
/gene="OR 13.1"
complement(<42101..43087)
/gene="OR 13.1"
/codon_start=1
/product="odorant receptor 13.1"
/protein_id="AAF27264.1"
/db_xref="GI:6739477"
/translation="MSAGNISFYKFEFFIVGPGIQLQGYFDSVAALMLFVFCILAGNG
TFLLIFLSLEKSLHPRVYFIINLVASOVLFSTTLPKIIISRYFQDGTISFLGCFVOM
YFVHFGTVNAFILAIFADRYVNAICNPLRYPNIVTKWIFYLCLISWIANVTPLMW
VIRAPLPYCAENTIIQCYCDHVSITSLACTNRAPYSIPAFVLWALLTPPLAFIVES
XCALILAVLRISSTQARLKTFTSCPOLIIIALIYFIPRCFIYLSNIGIYFFSTDRLA
IMMYSLPPMFPNPLIYCLRTKDVKECLVKRLKAINVNFNLTKVDISSLVSS"
complement(46118..46169)
/rpt_type=tandem
/rpt_unit=ta
complement(48998..48958)
/rpt_type=tandem
/rpt_unit=ca
complement(<50334..>51311)
/gene="OR 2.8"
/product="odorant receptor 2.8"
complement(<50334..>51311)
/gene="OR 2.8"
complement(50334..51311)
/gene="OR 2.8"
/codon_start=1
/product="odorant receptor 2.8"
/protein_id="AAF27265.1"
/db_xref="GI:6739478"
/translation="MSDSNSSFSSNISIVHPEYFFITGLSGTEYSNYYVIFLFFLIY
AVIGNFVLLIILALERSHSPKYIAVFNALADIGEMNALVPMNIMKELFESQYISYN
ACVMNMFVFFHFTMQSTPLVVMYSYDRFAICLPRYHAIVNNTVISVFLVLMAFNV
FVLLVYGVMITRISICKSNVLYCYDHPGVYKACNDNSINSLFKLSTALYLVAPL
IIIFLSYGVFLALSITTWEGRLKALKTCVAHLLLAGSFFVPIFISYIATSMNSLTP
NARVLSLSYVLPPLNPLIYLVNLTAEIKLKKIKNRSTKTVHVAS"
complement(<60772..>61752)
/gene="OR 2.11"
/product="odorant receptor 2.11"
complement(<60772..>61752)
/gene="OR 2.11"
complement(60772..61752)
/gene="OR 2.11"
/codon_start=1
/product="odorant receptor 2.11"
/protein_id="AAF27266.1"
/db_xref="GI:6739479"

```

```

Query Match
Best Local Similarity 3.08; Score 44.2; DB 8; Length 140658;
Matches 118; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
Qy 1243 atggtttatatgcaacggaagaattgcccattgttggaattgttttttagatcatcaaa 1302
Db 119779 ATGCTGTATATGCCAAAATATATATATATATATATATATATATATATATATAT 119720
Qy 1303 ggcctctacagattcttagggaatggttcaggcttttttagaagaattgctttattgc 1362
Db 119719 TAATCATATATATATGATTTTGAATTTTATGACAAATCTTATTGTCATATCTGTTTA 119660
Qy 1363 aacaggtagagaacataaacatagacagatgctatctgaagagataagctctctatgct 1422
Db 119659 TAATGTTCTTAATATATTTTTTATAAAGTTGTCTAAGCAGAACACAGCTAGTGAATGT 119600
Qy 1423 aaagaaatgacgcatgacataaaacaagcatcataaagattaaaaaataaaaaa 1482
Db 119599 GCCTATTTTAGGTGTAAATATACCAGAAAAACAAAATGGCAAAAAAATAAAAAA 119540

```


[illegible]

[illegible]

ACCESSION	AL078590
VERSION	AL078590.28
KEYWORDS	GI:8218068
SOURCE	HTG.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 195021)
JOURNAL	Babbage.A.
COMMENT	<p>Direct Submission</p> <p>Submitted (24-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1LS, UK. E-mail enquiries: humquery@sanger.ac.uk</p> <p>request: clonerequest@sanger.ac.uk</p> <p>On Jun 3, 2000 this sequence version replaced gi:7159744.</p> <p>During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.</p> <p>This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.</p> <p>The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6</p> <p>RP11-557H15 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/</p> <p>VECTOR: pBACe3.6</p>
FEATURES	<p>Location/Qualifiers</p> <p>1..195021</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/chromosome="6"</p> <p>/map="q23.1-23.3"</p> <p>/clone="RP11-557H15"</p> <p>/clone_lib="RPCI-11.2"</p> <p>3..514</p> <p>/note="match: GSS: Em:AQ434946"</p> <p>33..1572</p> <p>/note="match: STS: Em:AF192016"</p> <p>111..225</p> <p>/note="L2 repeat: matches 2617..2738 of consensus"</p> <p>493..803</p> <p>/note="AluX repeat: matches 1..311 of consensus"</p> <p>1973..2282</p> <p>/note="AluX repeat: matches 1..310 of consensus"</p> <p>2986..3236</p> <p>/note="L2 repeat: matches 2517..2750 of consensus"</p> <p>3609..3925</p> <p>/note="AluJo repeat: matches 1..311 of consensus"</p> <p>4328..4407</p> <p>/note="L2 repeat: matches 1890..1969 of consensus"</p> <p>4416..4537</p> <p>/note="61 copies 2 mer to 68% conserved"</p> <p>4538..4866</p> <p>/note="AluJo repeat: matches 1..310 of consensus"</p> <p>5092..5622</p> <p>/note="match: GSS: Em:AQ701736"</p> <p>5342..5518</p>

7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 16.23 Seconds
(without alignments)
1722.493 Million cell updates/sec

Title: US-09-545-072a-2
Perfect score: 1922
Sequence: 1 MEAYKQWVRNREVQSFSGS.....EKIVELLEGAQSRYYTISGS 367

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1773	92.2	360	2 T00882	hypothetical prote
2	99	5.2	1273	2 C96767	unknown protein F2
3	96	5.0	376	2 T40591	hypothetical prote
4	96	5.0	987	2 A64474	hypothetical prote
5	95.5	5.0	932	2 C70191	penicillin-binding
6	93	4.8	398	2 C83499	probable transglyc
7	93	4.8	1306	2 S25370	MSB2 protein - yea
8	91	4.7	296	2 E82250	conserved hypothet
9	90.5	4.7	1277	2 T30532	neural cell adhesi
10	90	4.7	627	2 T84693	probable protein k
11	90	4.7	639	2 T15168	hypothetical prote
12	89.5	4.7	1849	2 T00415	hypothetical prote
13	89	4.6	333	2 T51650	probable transcrip
14	89	4.6	549	2 B32372	male-specific doub
15	89	4.6	3165	2 S15010	hypothetical prote
16	88	4.6	1706	1 OYBRC	cyclolysin - Borde
17	86.5	4.5	313	2 T05256	hypothetical prote
18	86.5	4.5	879	2 E69792	conserved hypothet
19	86.5	4.5	1025	1 A57676	protein kinase Xa2
20	86.5	4.5	4466	1 S17231	dyein beta heavy
21	86	4.5	1872	2 T30888	vitellogenin - Ath
22	85.5	4.4	758	2 S61834	cheA protein - Rhi
23	85.5	4.4	784	2 H83158	Fe(III) dicitrate
24	85.5	4.4	879	2 T41059	replication licens
25	85.5	4.4	2380	2 E71604	hypothetical prote
26	85	4.4	248	2 T44405	methionyl aminopep
27	85	4.4	414	2 C72036	conserved hypothet
28	85	4.4	414	2 F86587	Cr605 hypothetical
29	85	4.4	535	2 T24309	hypothetical prote

30 85 4.4 1238 2 T32625 hypothetical prote
31 84 4.4 1406 2 T17429 gag-pol polypeptid
32 84 4.4 3411 1 GNMVY genome polypeptid
33 84 4.4 3411 1 GNMVY genome polypeptid
34 83.5 4.3 239 2 G71066 hypothetical prote
35 83.5 4.3 886 2 F83862 penicillin-binding
36 83.5 4.3 1325 1 S73723 probable lipoprote
37 83 4.3 202 2 C72011 SIN3 protein-bindi
38 83 4.3 513 2 S50915 probable heme expo
39 83 4.3 524 2 T09944 enamelin precursor
40 83 4.3 1142 2 T37455 enamelin precursor
41 82.5 4.3 2145 2 J04747 adenylate cyclase
42 82 4.3 401 2 T48495 hypothetical prote
43 82 4.3 566 2 S75233 ABC transporter sl
44 82 4.3 767 2 T41344 probable zinc-fing
45 82 4.3 2114 2 E96505 hypothetical prote

ALIGNMENTS

RESULT 1

T00882
hypothetical protein At2g45690 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F17K2.22
C:Species: Arabidopsis thaliana (mouse ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00882; F84893
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
submitted to the EMBL Data Library, March 1998
A:Description: Arabidopsis thaliana chromosome II BAC F17K2 genomic sequence.
A:Reference number: 214207
A:Accession: T00882
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-360 <R00>
A:Cross-references: EMBL:AC003680; NID:g2979540; PID:g2979558
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
cuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: F84893
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <STO>
A:Cross-references: GB:AF002093; NID:g6598429; PIDN:AAC28524.2; GSPDB:GN00139
C:Genetics:
A:Gene: F17K2.22; At2g45690
A:Map position: 2
A:Introns: 26/1; 99/2; 297/3; 320/2; 341/1
C:Superfamily: Arabidopsis thaliana hypothetical protein F17K2.22

Query Match 92.2%; Score 1773; DB 2; Length 360;
Best Local Similarity 96.9%; Pred. No. 9.3e-148;
Matches 343; Conservative 1; Mismatches 0; Indels 10; Gaps 1;

Qy 24 GLTWLLPKFSASEIGPEAVTAFLGIFTTNEHIENAPTRGHVSGSGNDPSLSYPLLI 83
|||||
Db 7 GLTWLLPKFSASEIGPEAVTAFLGIFTTNEHIENAPTRGHVSGSGNDPSLSYPLLI 66
|||||
Qy 84 ALLKDLFTVVEAAEHFGDKKNYIILTEAMKAVIRLALFRNSGYKMLGGTTPNEEK 143
|||||
Db 67 ALLKDLFTVVEAAEHFGDKKNYIILTEAMKAVIRLALFRNSGYKMLGGTTPNEEK 126
|||||
Qy 144 DSNQSESONRAGSGRNGLPHGLGNQNHHPNWNLEGRAMSALSSFGQNARTTTSSTPGWS 203
|||||
Db 127 DSNQSESONRAGSGRNGLPHGLGNQNHHPNWNLEGRAMSALSSFGQNARTTTSSTPGWS 186
|||||
Qy 204 RRIHQQAVIEPPMIKERRRTMSELLTEKGVNGALFAIGEVLYITRPLIYVLFIRKYGVR 263
|||||

Db 187 RRIHQQAIVLEPPMIKERRRNTSELITKGVNGALFAIGEVLYITRPLIYVLFIRKYGVR 246
Qy 264 SWIPWALSISVDTLGMGLANSKWGEKSKQVHFSQPKDE-----LRRKKLIWA 313
Db 247 SWIPWALSISVDTLGMGLANSKWGEKSKQVHFSQPKDEVRNLIDLNTTTRRRKLIWA 306
Qy 314 LYLMDRPFPTKYTRKLESOKKLELPLIGLFTKIVELLEGASQRYTYSIS 367
Db 307 LYLMDRPFPTKYTRKLESOKKLELPLIGLFTKIVELLEGASQRYTYSIS 360
RESULT 2
C96767
unknown protein F2P9.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96767
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96767
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1273 <STO>
A:Cross-references: GB:AE005173; NID:g7109476; PIDN:AAF36740.1; GSPDB:GN00141
C:Genetics:
A:Gene: F2P9.17
A:Map position: 1

Query Match 5.2%; Score 99; DB 2; Length 1273;
Best Local Similarity 19.7%; Pred. No. 3.7;
Matches 73; Conservative 43; Mismatches 105; Indels 150; Gaps 17;
Qy 40 PEAVTALGIFTTINEHIENATPRGHVSSGNDPSLSYPLLIALLKLE----- 90
Db 950 PEPVAAV-----DETQHL-----EPR---MESQNPQSHENPIVHEIPSDVEGTELAHR 996
Qy 91 -----TVVEVAAEHFVGDKKNYIILTEAMKAVIRLALFRNSGYKMLQG-- 135
Db 997 EANPPTKEPQEPDVSVSVSH-----EVKKSIVIRKV-RSGATSRAECSA 1042
Qy 136 -----GETPNEKSNQSESONRAGNSGRNIGPHGLGNQNHHPN-- 176
Db 1043 RTIERSQGIIVRHIDRGQTSASVAPQRISTDAV-----SISQNHVEEVNSC 1092
Qy 177 --LGRAMSALES-----FGONAR-TTTSSTQPSRHIOQAQVTEPPMIKERRT 224
Db 1093 HDVGSRTASIGSVKFASEGIGFKEGLQCTAESGKSTOKADNNRTVPPSLP----- 1147
Qy 225 MSELLTEKGVNGALFAIGEVLYITRPLIYVLFIRKYGVRSWIPWALSISVDTLGMGLAN 284
Db 1148 ----LDHSMENEAQ-----QKVA-----SLQTLSIG----- 1169
Qy 285 SKWGEKSKQVHFSQPKDELRRKLIWALYLMRDPEFTKYTRKLESOKKLELPLIG 344
Db 1170 ----KEKEK-----KDEKKEKKRR-----EDPVYLEKRLKKEKKKEKMAKLVS 1214
Qy 345 FLTEKIVELLE 355
Db 1215 STTDPAKKIE 1225
RESULT 3

T40591
hypothetical protein SPBC646.15c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C:Accession: T40591
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A:Reference number: 221938
A:Accession: T40591
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-376 <SEE>
A:Cross-references: EMBL:AL03216; PIDN:CAA22819.1; GSPDB:GN00067; SPDB:SPBC646.15c
A:Experimental source: strain 972h-; cosmid c646
C:Genetics:
A:Gene: SPDB:SPBC646.15c
A:Map position: 2
A:Introns: 49/1; 126/2; 312/2; 350/1
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC646.15c
Query Match 5.0%; Score 96; DB 2; Length 376;
Best Local Similarity 26.5%; Pred. No. 1.3;
Matches 39; Conservative 19; Mismatches 45; Indels 44; Gaps 8;
Qy 227 ELLTEKGVNGALFAIGEVLYITRPLIYVLF-----RK-----YGVRSWIPWALSISVDTL 277
Db 207 EIFTKRLPLNRIFS--NFIKVCRLIYLMFMWHRKQKSSSLKVRPWPWPVAFVFEVI 264
Qy 278 -----GMCILANSKWGEKSKQVHF-----SGPEKDELRR---RKL 311
Db 265 SOLIDRRCESATSSRQGFLEERT-----NQSQFHVVVMAFTQGRFYDEFTKHWNRSL 320
Qy 312 WALYLMRDPFFTKYTRKLESOKKLE 338
Db 321 WWSNI---PVFGKYLILLSVEERQKSL 344
RESULT 4
A64474
hypothetical protein MJ1394 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: A64474
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: A64474
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-987 <BUI>
A:Cross-references: GB:U67579; GB:L77117; NID:g1592037; PIDN:AAB99404.1; PID:g1592041
C:Genetics:
A:Map position: FOR1343894-1346857
A:Start codon: GTG
Query Match 5.0%; Score 96; DB 2; Length 987;
Best Local Similarity 20.4%; Pred. No. 4.8;
Matches 60; Conservative 37; Mismatches 131; Indels 66; Gaps 10;
Qy 100 FYGDKKNYIILTEAMKAVIRLALFRNSGYKMLLOGGETPNEEKDSNSESQ----- 152
Db 577 FYGDKKNYIILTEAMKAVIRLALFRNSGYKMLLOGGETPNEEKDSNSESQ-----KDYNTSLNAINWTNIT 624
Qy 153 -RAGNSGRNIGPHGLGNQNHHPNWNLEGRAMSALSSFGONARTTTSSTPG-----WSRR 205
Db 625 LNWNSDSATLVFNVLGNYSYERDNLAKYGFKILFVNGTNTNTSIKGVYASGSYS 684

QY 206 IQHQ-----QAVIEPPIKERRRTMSELLTEKGVNGALFAIGEVLYITRPLIYVLFIRKY 260
Db 685 TDGTTGTEININWTFKNDKASYFNLTLNI-----WAVNKSAYELYWNPF 733
QY 261 GYRSWIP---WAISLSVDTLGMLLANSKWGKSKQVHFSQPEKDELRRKKLIWA---L 314
Db 734 NKSIMDSNVTITPNID-----IPGVEVNSKTYNFTFSGV-----PIYVANCSE 779
QY 315 YLMRDPFFTKYTRQKLESSQKKLELPLIGF---LTKIVELLEGAQSRVYTI 364
Db 780 TLSKDYILLNEVSIQGSVVYVEIYVVGSLYKIKVHPDADGIYDIYIV 833
RESULT 5
C70191
penicillin-binding protein (pbp-3) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: C70191
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White-
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: C70191
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-932 <KLE>
A:Cross-references: GB:AE001173; GB:AE000783; NID:g2688665; PIDN:AAAC67082.1; PID:g268867
A:Experimental source: strain B31
Query Match 5.0%; Score 95.5; DB 2; Length 932;
Best Local Similarity 21.18; Pred. No. 4.9;
Matches 84; Conservative 39; Mismatches 121; Indels 155; Gaps 19;
QY 45 AFLG---IFTTINEHIENAPTP-----RGHVGSNGN----- 73
Db 393 AILGIDKIDKATKEYIKNKLTLPKLIQAPEGAMTIDTTSCAIRAMVGGSGHTKDNFNR 452
QY 74 -----DPSLSY-PLIIAILKDLTEV-----EVAEHFYGDKKNYI 109
Db 453 ATQAKVQPSAFKALYFAAAILDLKITAATMFSDSPVAFLNKNGEYVAPNGYG-KWRGN 511
QY 110 ILTE-----AMKAVIRLALFRNSGYKMLQGGTPE-EKD-----SN 146
Db 512 VITRALALSINIPALRILDRGFDLSAISYSSKLLGITDPKEIEKTPKVPYPLALGVISV 571
QY 147 QSESNRA-----GNSGRNLGPHGLGNQHNHPNWNLEGRAMSALSSFGQNARTTTSSTPGW 202
Db 572 SPIQMARAFAILNGSSEIEPYGIR-----YIEDRA-----GRIITNEEASILAK 616
QY 203 SRRIHQQAVIEP-----PMIKERRRTMSELLTE-KGVNGAL----- 238
Db 617 INKKEHOTQIVSPQATYIITDMKSTIQYGTLANQRYTNLKNFKSDIAGSGTTQNNWADG 676
QY 239 FAIGEVLYITRPLIYVLFIRKGVRSWIPWAIISLVDTLGMLLANSKWGKSKQVHFS 298
Db 677 WAIGVSPYIT-TAFVWGFDKK-----CYSLSGSTGTGLAGPS--MGEFMAEYHK 724
QY 299 GPEK-----DELRRKKLIWALYL 316
Db 725 LPKKVFKVPKAGIISIPVQAETGLLPEETADEKIINELFI 763
RESULT 6
C83499
probable transglycolase Pall171 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83499

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: C83499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <STO>
A:Cross-references: GB:AE004547; GB:AE004091; NID:g9947089; PIDN:AG04560.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1171
Query Match 4.8%; Score 93; DB 2; Length 398;
Best Local Similarity 26.8%; Pred. No. 2.5;
Matches 62; Conservative 26; Mismatches 85; Indels 58; Gaps 14;
QY 155 GNSGRNL---GPHGLGNQHH-----NPNWLEGR--AMSALSSFGQNARTTTSSTP 200
Db 198 GDGRDLVGSVPDALGSTANYLKKAGWRTGCPWGYEVKVPADFPASLAGRGRQPLSA-- 255
QY 201 GW-----SRRIHQQAVIEPPIKERRRTMSELLTEKGVNGALFAIGEVLYITRPLIYVLF 256
Db 256 -WARGVRRVVDGQPL-----PGGDEK-----AAILLPAGAQPAP-----LVRYNYDAIV--- 299
QY 257 IRKYGVRWSWIPWAIISLVDTL--GMGLLANSKWGKSKQVHFSQPEKDELRRKKLIWAL 314
Db 300 --SYNAEYALAIALLSDKRLGSGSLVAS---WPT-----DDPGISRLERKKLQKAL 347
QY 315 YLMR-----DPFTKTYTRQKLESSQKKLELPLIGFTEKIVELLEGAQ 358
Db 348 -LARGYDIGEADGLIGTSTKKAQAEKRLGLTPADGRAGRKILEAKGAQ 397
RESULT 7
S25370
MSB2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G4017; protein YGR014W
C:Species: Saccharomyces cerevisiae
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C:Accession: S25370; S64305
R:Bender, A.; Pringle, J.R.
Yeast 8, 315-323, 1992
A:Title: A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.
A:Reference number: S25370; MUID:92383951
A:Accession: S25370
A:Molecule type: DNA
A:Residues: 1-1306 <BEN>
A:Cross-references: GB:M77354; NID:g171993; PIDN:AAA34798.1; PID:g171994
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64071
A:Accession: S64305
A:Molecule type: DNA
A:Residues: 1-1306 <RIE>
A:Cross-references: EMBL:272799; NID:g1322977; PID:g1322978; MIPS:YGR014W
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:MSB2
A:Cross-references: SGD:S0003246; MIPS:YGR014W
A:Map position: 7R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
C:Keywords: transmembrane protein
F:3-19/Domain: transmembrane #status predicted <TM1>
F:1189-1205/Domain: transmembrane #status predicted <TM2>
Query Match 4.8%; Score 93; DB 2; Length 1306;
Best Local Similarity 21.6%; Pred. No. 13;
Matches 61; Conservative 49; Mismatches 83; Indels 90; Gaps 15;

Db 152 NDIDNHVGI-----DLNGVTSIASAPAAAYFDDREAKNISLRASGKP 194

QY 119 IRLAFNSCYKMLQGGTTPNEKDSNQESQNRAGSNRNLGPHGLGNQHNHPNLE 178

Db 195 VRWIEYNATFNLVNTLAPDRKPSIPLLS-----RKWNLGIFSQEHH----- 240

QY 179 GRAMSALSSFGONARTTSS--TPGWSRRIOHQQA---VIEPPMIKERRRMTS-----EL 228

Db 241 -----VGFSASTGVVASSHFFVLGWSFNIEGKESDFDIKLSLPPPTLSPSPPV 293

QY 229 LTKGVNGALFAIGEVYITRPLIYVLFIRKYGVRS-----WIPWALSVDTL--GMGLL 282

Db 294 STEKKSNTMLIIIVAAATVALMILIFSGFWFLRRDKIFFIGGARKFSYQTISNATGGF 353

QY 283 ANSKWGEKSKQVHFG 299

Db 354 DNSKLLGERNSGGSFYKG 370

RESULT 11

T15168

hypothetical protein F19B10.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15168

R:Gattung, S.

submitted to the EMBL Data Library, April 1997

A:Description: The sequence of C. elegans cosmid F19B10.

A:Reference number: Z18302

A:Accession: T15168

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-639 <GAT>

A:Cross-references: EMBL:AF000261; NID:g1947114; PID:g1947124; PIDN:AA52930.1; GSPDB:CN

A:Experimental source: strain Bristol N2; clone F19B10

C:Genetics:

A:Gene: CESP:F19B10.10

A:Map position: 2

A:Introns: 29/3; 67/1; 101/3; 137/3; 272/3; 339/3; 407/3; 509/3

Query Match 4.7%; Score 90; DB 2; Length 639;

Best Local Similarity 20.5%; Pred. No. 8.8;

Matches 53; Conservative 47; Mismatches 97; Indels 62; Gaps 12;

QY 1 MEAYKQWNRNREIVQSGFANGLTWLLPEKFSASEIG----PEAVTAL-----GIFT 51

Db 371 LEHLRDLFFGSRVFSQSPTRSMFNGYVISMNEIFNSNAQGIYFLEESIEVEKISSGLFG 430

QY 52 TINPHIENAPTRGH-VGSSGNDPSLSYPLLIALKDLTVVEVAAEHFYGDKKNYI- 109

Db 431 TPESKTIRDVP-PKGFTVGDVHSE--LKYLGLLEFFFNISYAEIAYKVVAEKKMKILR 487

QY 110 --ILTEAMKAVIRALFRNGYKML-----LOGGETPNEEKDSNQESQNRAG 155

Db 488 TCDMFDAVEICMLAVF--NGYPKLAQFLHSQHACHRIQGLQCKKCELEK----- 536

QY 156 NSGRNLGPHGLGNQHNHPNWL-----EGRAMSALSSFGONARTTSSTPGWSRRIOHQ 209

Db 537 -----HPW-----WNSWLYPMEVEDEKVDSTSSDDEEKKRTSTSELAGNRKKEQK 583

QY 210 QAVIEPP-----MIKERRT 224

Db 584 KLVAKDPEPSTMEKTRKS 602

RESULT 12

T00415

hypothetical protein H_248015.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 05-Nov-1999

C:Accession: T00415

R:Fulton, B.; Nhan, M.; Hawkins, J.; Beck, C.

submitted to the EMBL Data Library, September 1998

A:Description: The sequence of Homo sapiens PAC clone 248015.

A:Reference number: Z14148

A:Accession: T00415

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1849 <FUL>

A:Cross-references: EMBL:AC002483; NID:g3598729; PIDN:AAC35295.1; PID:g2341020

C:Genetics:

A:Map position: 13q12-q13

A:Note: H_248015.1

Query Match 4.7%; Score 89.5; DB 2; Length 1849;

Best Local Similarity 23.3%; Pred. No. 42;

Matches 48; Conservative 32; Mismatches 87; Indels 39; Gaps 10;

QY 76 SLSYPLLIAILKD-LETVEVAAEHFYGDKKWNYIIITEAMKAVIRLA-----LFRN 126

Db 1280 ALKQPLSAHALSDLLSRLVEVIGEH--GDEIQGYV-----MEALLTLEAAVDNLSCLKN 1332

QY 127 SGYKMLQGGTTPNEEKDSNQESQNRAG----NSGRNLGPHGLGNQHNHPNLEGRAM 182

Db 1333 SLLTVLSRSSPDLSSSSKLTASRKSTGQLNMNPGTSGNTATAERSHQ-----RSF 1386

QY 183 SALSSFGONARTTSTPGWSRRIOHQQAIVTEPPMIKERRRMTS--ELLTEKG-VNGALF 239

Db 1387 SVPKKFGVIDR---SSDPPRSATLDRIQACTQOGLSKSRSSSLKSLDTPSHINPTN 1443

QY 240 AIGEVLYITRPLI-----YVLFIR 258

Db 1444 LLATIFWTVTALMESDFEFYLMALR 1469

RESULT 13

T51650

probable transcription factor MYB36 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 06-Oct-2000

C:Accession: T51650

R:Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Kranz, H.D.; Denekamp, M.; Greco, R.

; Paz-Ares, J.; Weissshaar, B.

Plant J. 16, 263-276, 1998

A:Title: Towards functional characterisation of the members of the R2R3-MYB gene from

A:Reference number: Z14349; MUID:9839469

A:Accession: T51650

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-333 <KRA>

A:Cross-references: EMBL:AF062878; PIDN:AAC83600.1

C:Genetics:

A:Gene: MYB36

A:Map position: 5

C:Superfamily: myb-related protein MIXTA; myb DNA-binding repeat homology

Query Match 4.6%; Score 89; DB 2; Length 333;

Best Local Similarity 17.3%; Pred. No. 4.4;

Matches 59; Conservative 35; Mismatches 113; Indels 134; Gaps 12;

QY 12 REYVQSGFSGFANGTULW-LPEKFSASEIGPEAVTAFGLFIFTINEHIENAPTR-CHVG 69

Db 26 KDYIDKYGTGGN---WIALPOKIGLKRCKSCRLRWL-----NYLRPNIKHGG 70

QY 70 SSGNDPSLSYPLLIALKDLTVVEVAAEHFYGDKKWNYIILTEAMKAVIRALFRNSGY 129

Db 71 FSEEDRIILSLYLSI-----GSRWSIIAAQLPGRTDNDIKNYWNTKL 113

QY 130 KMLQGGTTPNEEKDSNQESQNRAGSNRNLGPHGLGN-----QHNHNPNWLEGRAMS 183

Db 114 KKKLLGRQKQNRQDSITDSTENNLSNNNNKSPQNLNSALEKILQLHMQNLQ-----S 169

QY 184 ALSFGQN-----ARTTSSIPGWSRR-----IQHQAVIEPPMIKERRR 223

Db 170 PFSSFYNNPILWPKLHPLLOSTTTTQNPKLASQSFHPLGVNDHQH---NNTKLAQINN 226
Qy 224 TMSSELLTE-----KGVNGALFAIGEVLVITRP 250
Db 227 GASSLYSENVSQNPAPHEQPNFGFSQDLRLDNHNMDFMNRGVSKELFQVGNFEFLTN- 285
Qy 251 LIYVLFIRKYGVRSWIPWAISSLVSDTLGMLLANSKMWGEK 291
Db 286 -----GSSWSEE 293

RESULT 14
B32372
male-specific doublesex protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 02-Nov-1989 #sequence_revision 02-Nov-1989 #text_change 16-Feb-1997
C:Accession: B32372
R:Burtis, K.C.; Baker, B.S.
Cell 56, 997-1010, 1989
A:Title: Drosophila doublesex gene controls somatic sexual differentiation by producing
A:Reference number: A32372; MUID:89168451
A:Accession: B32372
A>Status: preliminary
A:Molecule type: DNA: mRNA
A:Residues: 1-549 <BUR>
A:Cross-references: GB:M25294; GB:J03156
C:Genetics:
A:Gene: FlyBase:dsx
A:Cross-references: FlyBase:FBgn0000504
C:Keywords: alternative splicing; DNA binding; transcription regulation

Query Match 4.6%; Score 89; DB 2; Length 549;
Best Local Similarity 23.5%; Pred. No. 8.8;
Matches 51; Conservative 16; Mismatches 78; Indels 72; Gaps 7;

Qy 45 AFLGIETTNEHIENAPTRGHVSGSNDPSLSY-----PLLIATLKDL 89
Db 333 AILPISVSVRKNGANVPL-----GQVFLDYCKLEKFRYPWELMPLMYILKDA 384
Qy 90 ETVVEAAEHFYGDKNWYIILTEAMKAVIRLA--LFRNSGYKMLQGGT-----PNEE 142
Db 385 DANIEEASRR-----IEEARVEINRTVAQIYNYITPMALVNGAPWYLTYPSE 433
Qy 143 KDSNQSESONRAGNSGNL-----GPHGLGNQNHNNPNWLEG 179
Db 434 -----QGRYGAHFTHLPTQICPTPEPLALSRSPSPSGPSAVHNQKPSRPGSSNG 485
Qy 180 RAMSALSSFGQNARTTTSTPGWRSRIHQHQAIEPP 216
Db 486 TVHSAASPTWTTMTATTSSTPTLSRRQRSATPTTP 522

RESULT 15
S15010
hypothetical protein B - Cryphonectria hypovirus 1
C:Species: Cryphonectria hypovirus 1
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C:Accession: S15010
R:Shapira, R.; Choi, G.H.; Nuss, D.L.
EMBO J. 10, 731-739, 1991
A:Title: Virus-like genetic organization and expression strategy for a double-stranded R
A:Reference number: S15009; MUID:91184117
A:Accession: S15010
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-3165 <EMB>
A:Cross-references: GB:M57938; NID:g331157; PIDN:AAA67458.1; PID:g331159

Query Match 4.6%; Score 89; DB 2; Length 3165;
Best Local Similarity 20.3%; Pred. No. 97;

Matches 82; Conservative 49; Mismatches 149; Indels 124; Gaps 18;
Qy 4 YKQW--VWNRREVVSQSFSGFANGLTWLLPEKFSASEIGPEAVTAF-----LGITTT 52
Db 1757 FQWVEVFADRDIM-----LPKGVELYIKKYSA--GTPFISSFYKSKALKQAGVMDV 1808
Qy 53 INEHIENAPTRGHVSGSNDPSLSY-----PLLIATLKDLTETVVEAAEHF 100
Db 1809 IRKNALECI-----STGKYPTQFYHAFKSAQVPGOPLLAPRMKDLRTVVSSELSAY 1860
Qy 101 YGDKKNWYIILTEAMKAVIRLALFRNSGYKMLQGGTPENEKDSNQSESQN-----RAGN 156
Db 1861 MYDQ-----IFQTEANKRITWETYGAGSGMPLSQSMARIWDELHLRKREGQFIITADATA 1916
Qy 157 SGRNLGP---HG-----LGNQNHNNPNWNLGRAMSALSFG-----QNARTTTSSTPGWS 203
Db 1917 YDSNCKPALFHGAGKLIVELGFQNHPSG---KGRQPAQVVQCKFEAMQANWVMGITEPSYT 1973
Qy 204 RRIHQQAVIEPPMIKERR-----RTMSELLTEKGVNGALFAIGEVLVITRPLIYVLFIRK 259
Db 1974 ALTFHVPDVAVRHELESKYPAHFATFSELLAHNVN-----VTE 2012
Qy 260 YGVRSW-----IPWAISSLVSDTLGMLLANSKMWGEKSKQVHFSQPEKDELRR 307
Db 2013 WKRLSWEERKACARDMQAVPGVKVFLTNDPALR--LOGSSWQGSFTTE-----PKRDE--- 2062
Qy 308 RKLWALYLMRDPFFTKYTRQKLESSQKKLELIPLIGLITKIV 351
Db 2063 -----FRKYOTFYDYSKAAMREDIKRIVFANREVI 2092

Search completed: July 11, 2001, 11:35:55
Job time: 54 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 24.54 Seconds
(without alignments)
1978.645 Million cell updates/sec

Title: US-09-545-072A-2
Perfect score: 1922
Sequence: 1 MEAYKQWNRNREYVQSGFS.....EKIVELLEGAQSRVYIISGS 367

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_unclassified.*
- 13: sp_invertebrate.*
- 14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	100.0	367	10 Q9XEG0	Q9xeg0 arabidopsis
2	1773	92.2	360	10 Q9SLI9	Q9sl19 arabidopsis
3	249.5	13.0	336	4 Q9V5Y5	Q9v5y5 homo sapien
4	227.5	11.8	341	5 Q9VPB9	Q9vpb9 drosophila
5	97	5.0	855	10 Q9FYN9	Q9fyn9 oryza sativ
6	96.5	5.0	1957	4 O15064	O15064 homo sapien
7	96	5.0	376	3 Q94516	Q94516 schizosacch
8	95.5	5.0	932	2 O51674	O51674 borrelia bu
9	94	4.9	1005	4 O75336	O75336 homo sapien
10	93	4.8	398	2 O914G6	O914g6 pseudomonas
11	92	4.8	1196	3 Q9HDP9	Q9hdp9 cephalospor
12	91	4.7	296	2 Q9KT82	Q9kt82 vibrio chol
13	90.5	4.7	330	10 Q9M8K2	Q9m8k2 arabidopsis
14	90.5	4.7	1277	13 Q98902	Q98902 fugu rubrip
15	90	4.7	627	10 Q9ZW09	Q9zw09 arabidopsis
16	90	4.7	639	5 O02075	O02075 caenorhabdi
17	89.5	4.7	1257	2 P96746	P96746 corynebacte
18	89.5	4.7	1295	2 O34110	O34110 helicobacte
19	89.5	4.7	1849	4 O14572	O14572 homo sapien

20	89.5	4.7	3012	4 Q9Y3N6	Q9y3n6 homo sapien
21	89	4.6	333	10 Q9ZTE6	Q9zte6 arabidopsis
22	89	4.6	333	10 Q9FKL2	Q9fkl2 arabidopsis
23	89	4.6	3165	14 Q04350	Q04350 cryphonectr
24	88.5	4.6	886	11 Q9ES60	Q9es60 mus musculu
25	88.5	4.6	890	11 Q9ES59	Q9es59 mus musculu
26	88.5	4.6	1072	11 Q9R243	Q9r243 mus musculu
27	88.5	4.6	1172	11 Q9R244	Q9r244 mus musculu
28	88	4.6	390	2 P70928	P70928 bacillus li
29	88	4.6	425	10 Q9ST79	Q9st79 oryza sativ
30	87.5	4.6	320	2 Q9FAF7	Q9faf7 agrobacteri
31	86.5	4.5	285	10 Q9FT68	Q9ft68 arabidopsis
32	86.5	4.5	313	10 Q9SVM6	Q9svm6 arabidopsis
33	86.5	4.5	879	2 O31504	O31504 bacillus su
34	86.5	4.5	1025	10 Q40640	Q40640 oryza sativ
35	86	4.5	304	2 Q9REA3	Q9rea3 enterococcu
36	86	4.5	1872	5 O17083	O17083 athalia ros
37	85.5	4.4	649	4 Q9P215	Q9p215 homo sapien
38	85.5	4.4	784	2 Q9HXB2	Q9hxb2 pseudomonas
39	85.5	4.4	827	4 Q9ULJ0	Q9ulj0 homo sapien
40	85.5	4.4	881	3 Q9UVF5	Q9uvf5 yarrowia li
41	85.5	4.4	1290	2 O34111	O34111 helicobacte
42	85.5	4.4	2380	5 Q96266	Q96266 plasmodium
43	85	4.4	414	2 Q9Z7D0	Q9z7d0 chlamydia p
44	85	4.4	535	5 O18009	O18009 caenorhabdi
45	85	4.4	1238	5 O61198	O61198 caenorhabdi

ALIGNMENTS

RESULT 1
Q9XEG0 PRELIMINARY; PRT; 367 AA.
AC Q9XEG0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SHRUNKEN SEED PROTEIN.
DE SSEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RX MEDLINE=99212146; PubMed=10195899;
RA Lin Y., Sun L., Nguyen L.V., Rachubinski R.A., Goodman H.M.;
RT "The Pex16p homolog SSE1 and storage organelle formation in
RT Arabidopsis seeds";
RL Science 284:328-330(1999).
DR EMBL; AF085354; AAD30661.1; -.
DR InterPro; IPR000002; -.
DR ProDom; PD004563; -; 1.
SQ SEQUENCE 367 AA; 41624 MW; 35EE727140C485B8 CRC64;

Query Match 100.0%; Score 1922; DB 10; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.6e-166;
Matches 367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MEAYKQWNRNREYVQSGFSANGLTWLLPEKFSASEIGPEAVTAFIGITTTNEHIEN	60
Db	1	MEAYKQWNRNREYVQSGFSANGLTWLLPEKFSASEIGPEAVTAFIGITTTNEHIEN	60
QY	61	APTPRGHVSSGNDPSLSYPLLIILKLETVVEVAEAEHFYGDKNWYIILTEAMKAVIR	120
Db	61	APTPRGHVSSGNDPSLSYPLLIILKLETVVEVAEAEHFYGDKNWYIILTEAMKAVIR	120
QY	121	LALFRNSGYKMLLOGETPNEEKDSNQESQNRAGNSGRNLGPHGLGNQHNHPNLEGR	180
Db	121	LALFRNSGYKMLLOGETPNEEKDSNQESQNRAGNSGRNLGPHGLGNQHNHPNLEGR	180

Db 121 LALFRNSGYKMLQGGTPEEKDSNQSNRAGNSGRNLGPHGLGNHNPWNLEGR 180
QY 181 AMSALSSFGQNARTTSSPGWSRRIOHQOAVIEPPMIKERRRTMSELLTEKGVNGALFA 240
Db 181 AMSALSSFGQNARTTSSPGWSRRIOHQOAVIEPPMIKERRRTMSELLTEKGVNGALFA 240
QY 241 IGEVLYITRPLIYVLFIRKYGVRSPWPAISLSVDTLGMGLLANSKWKGEKSKQVHFSGP 300
Db 241 IGEVLYITRPLIYVLFIRKYGVRSPWPAISLSVDTLGMGLLANSKWKGEKSKQVHFSGP 300
QY 301 EKDELRRRLKIALYLMRDPFTTKYTRQKLESSQKLELIPLIGLFTKEIVLELLEGAQSR 360
Db 301 EKDELRRRLKIALYLMRDPFTTKYTRQKLESSQKLELIPLIGLFTKEIVLELLEGAQSR 360
QY 361 YTYISGS 367
Db 361 YTYISGS 367
RESULT 2
Q9SLL9 PRELIMINARY; PRT; 360 AA.
AC Q9SLL9
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE AT2G45690 PROTEIN.
GN AT2G45690.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.B., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana";
RL Nature 402:761-768(1999).
DR EMBL; AC004665; AAC28524.2; -;
SQ SEQUENCE 360 AA; 40603 MW; 79EE1446FFDADE41 CRC64;

Query Match 92.2%; Score 1773; DB 10; Length 360;
Best Local Similarity 96.9%; Pred. No. 5.7e-153;
Matches 343; Conservative 1; Mismatches 0; Indels 10; Gaps 1;
QY 24 GLTWLLPEKFSASEIGPEAVTAFLGIFTTNEHIENAPTGRGHVSSGNDPSLSYPLLI 83
Db 7 GLTWLLPEKFSASEIGPEAVTAFLGIFTTNEHIENAPTGRGHVSSGNDPSLSYPLLI 66
QY 84 AILKDLTVEVAEAEHFYGDKKWNIITTEAMKAVIRLALFRNSGYKMLQGGTPEEK 143
Db 67 AILKDLTVEVAEAEHFYGDKKWNIITTEAMKAVIRLALFRNSGYKMLQGGTPEEK 126
QY 144 DSNQSNRAGNSGRNLGPHGLGNHNPWNLEGRAMSALSSFGQNARTTSSPGWS 203
Db 127 DSNQSNRAGNSGRNLGPHGLGNHNPWNLEGRAMSALSSFGQNARTTSSPGWS 186
QY 204 RRIHQOQAVIEPPMIKERRRTMSELLTEKGVNGALFAIGEVLITRPLIYVLFIRKYGVR 263
Db 187 RRIHQOQAVIEPPMIKERRRTMSELLTEKGVNGALFAIGEVLITRPLIYVLFIRKYGVR 246
QY 264 SWIPWPAISLSVDTLGMGLLANSKWKGEKSKQVHFSGPEKDE-----LRRKKLIWA 313
Db 247 SWIPWPAISLSVDTLGMGLLANSKWKGEKSKQVHFSGPEKDEVRNLLDNTLRRKKLIWA 306

QY 314 LYLMRDPFTTKYTRQKLESSQKLELIPLIGLFTKEIVLELLEGAQSRITYISGS 367
Db 307 LYLMRDPFTTKYTRQKLESSQKLELIPLIGLFTKEIVLELLEGAQSRITYISGS 360
RESULT 3
Q9Y5Y5 PRELIMINARY; PRT; 336 AA.
AC Q9Y5Y5
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE PEROXISOMAL BIOGENESIS FACTOR 16.
GN PEX16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99122987; PubMed=9922452;
RA South S.T., Gould S.J.;
RT "Peroxisome synthesis in the absence of preexisting peroxisomes";
RL J. Cell Biol. 144:255-266(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Honsho M., Tamura S., Shimozawa N., Suzuki Y., Kondo N., Fujiki Y.;
RT "Mutation in PEX16 is causal in the Peroxisome-deficient Zellweger
RT Syndrome of Complementatation Group D.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF118240; AAD22466.1; -;
DR EMBL; AB016531; BAA88826.1; -;
SQ SEQUENCE 336 AA; 38643 MW; 0C0688C52E7AE47C CRC64;

Query Match 13.0%; Score 249.5; DB 4; Length 336;
Best Local Similarity 25.4%; Pred. No. 1.8e-14;
Matches 94; Conservative 68; Mismatches 157; Indels 51; Gaps 12;
QY 4 YKQWVRNREYVQSGFANGLTWLLPEKFSASEIGPEAVTAFLGIFTTNEHIENAPT 63
Db 11 YQEVTRPATAQLETAVRGFSVLLAGRADSHSELSELVYSASNLVLVNDGILRKELR 70
QY 64 PRGHVSSGNDPSLSYPLLI-----TAILKDLTVEVAEAEHFYGD-KKWNITLTEAMKAVI 119
Db 71 KKLPPV-----SLSQKLLTWSLVECEVEFEMGAAKVMGEVGRWLVTALQLAKAVL 123
QY 120 RLALFRNSGYKMLQGGTPEEKDSNQSNRAGNSGRNLGPHGLGNHNPWNLEGR 179
Db 124 RMLLL--LWFKAGLIQ--TSPPIVPLDRETQAQPPDGD-----HSPGNHE- 163
QY 180 RAMSALSSFGQNARTTSSPGWSRRIOHQOAVIEPPMIKERRRTM--SELLTEKGVNGA 237
Db 164 --QSYGKRNRVVRVLTQNTP--SLHSRWGA----PQREGQQQOHHSELSATPTPLGL 215
QY 238 LFAIGEVLITRPLIYVLFIRKYGVRSPWPAISLSVDTLGMGLLANSKWKGEKSKQVHF 237
Db 216 QETAEFLYIARPLHLGLGLGQSRKSWKPLLAVVDVTSLSLSDRKGLTTR----- 269
QY 298 SGPEKDELRRLKIALYLMRDPFTTKYTRQK-LESSQKLELIPLIGLFTKEIVLELLE 356
Db 270 ---ERRELRRTILLLYLLRSFPYDRFSEARILFLIQLLADHPGVGLVTRPLMDYLP 326
QY 357 AOSRYTYISG 366
Db 327 WQKIYFYSWG 336

RESULT 4
Q9VPB9 PRELIMINARY; PRT; 341 AA.
ID Q9VPB9
AC Q9VPB9;

RESULT 6
O15064
ID O15064 PRELIMINARY; PRT; 1957 AA.
AC O15064;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE KIAA0357 PROTEIN (FRAGMENT).
GN KIAA0357
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002355; BAA21573.1; -.
DR InterPro; IPR000901; -.
DR InterPro; IPR001993; -.
DR PROSITE; PS00867; CPASE_2; UNKNOWN_1.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_2.
FT NON_TER 1
SQ SEQUENCE 1957 AA; 222295 MW; 6F20BA68F4F153B2 CRC64;

Query Match 5.0%; Score 96.5; DB 4; Length 1957;
Best Local Similarity 21.1%; Pred. No. 19;
Matches 74; Conservative 44; Mismatches 136; Indels 97; Gaps 17;
QY 56 HITENAPTGRH-----VGSSNDPSLSYPLLIATLKDETVVEVAABHFY----- 101
DB 356 HINRIEIPRGALLVGVGSGCKQ---SLTRLAFFISSMD-VFQITLRKGVIQDFKMDL 411
QY 102 -----GKKNNYL-ILTEAMKAVIRIALFRNSGYKMLLOGETTPNEEKSNQESQNR 153
DB 412 ASICLAGVKNLTVPMTDAQVADERFLVIND----LLASGEIPDLYSD--DEVENI 464
QY 154 AGNSGRNLGPHGIGNONHNPNL-----EGRAMSALSSFGQNARTTTSSTPG----- 201
DB 465 ISNVRNEVSQGLVD--NRENCWFFIDIRRLQKLVTLFCFSPVGNKLRVSRKFFPAIVNCT 523
QY 202 -----W-----SRRIHQQAVIEPPMTKERRRTMSELLTEKGVNGALFAIGEVLV 246
DB 524 AIHWFHEWQQALESVSLRFLQNTGIEPTVKVQISKFMFAVHTSVNOTSQSYLSNEQRY 583
QY 247 -ITRPLIYVLFTRKGVRSWI---PWALSLVDITLGMGLLANSKMWGEKSKQVHPSGPEK 302
DB 584 NYTPKSFLEFIRLY-QSLLRHRKELCKKTERLENGLL-----KLHSTSAQV 630
QY 303 DELRRRLKIWALYLRMDPPFTYTRQKLESSOKLE---LIPLIGFLPEKI 350
DB 631 DDLK-----AKLAQAEVKLNKEDAKLQVGVGVEDTKV 665

RESULT 7
O94516
ID O94516 PRELIMINARY; PRT; 376 AA.
AC O94516;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE SPBC646.15C.
GN SPBC646.15C.
OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035216; CAA22819.1; -.
DR InterPro; IPR003015; -.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 44540 MW; A2221F31E5814692 CRC64;
Query Match 5.0%; Score 96; DB 3; Length 376;
Best Local Similarity 26.5%; Pred. No. 1.9;
Matches 39; Conservative 19; Mismatches 45; Indels 44; Gaps 8;
QY 227 ELLTEKGVNGALFAIGEVLITRPLIYVLFY----RK-----YGVRSWIPWALSLSVDTL 277
DB 207 EITRRLPNLRIFS--NFIKVCRPLIYMLFMWHWKRKQKSSSLKVRPGWPIVAFVEVI 264
QY 278 -----GMGLLANSKMWGEKSKQVHF-----SGPKEDELRR-----RKLI 311
DB 265 SQLIDRRCESATSSRQGFGLERRT---NQSQFHFVVAFTQGRFYDEFTKHWINSLS 320
QY 312 WALYLMRDPPTKYTRKLESSQKKLE 338
DB 321 WVNSI---PVFGKYLLLSVEERQKSLE 344
RESULT 8
O51674
ID O51674 PRELIMINARY; PRT; 932 AA.
AC O51674;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PENICILLIN-BINDING PROTEIN (PBP-3).
GN BB0732.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uitterback T., Matthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL; AE001173; AAC67082.1; -.
DR TIGR; BB0732; -.
DR InterPro; IPR001264; -.
DR InterPro; IPR001460; -.
DR Pfam; PF00905; Transpeptidase; 1.
DR Pfam; PF00912; Transglycosyl; 1.
DR Prodom; PD001895; -; 1.
SQ SEQUENCE 932 AA; 105254 MW; 9D44BE34A10B8E77 CRC64;
Query Match 5.0%; Score 95.5; DB 2; Length 932;
Best Local Similarity 21.1%; Pred. No. 8;
Matches 84; Conservative 39; Mismatches 121; Indels 155; Gaps 19;

[illegible]

ID Q9802 PRELIMINARY; PRT; 1277 AA.
 AC Q9802;
 DT 01-FEB-1997 (TREMREL. 02, Created)
 DT 01-FEB-1997 (TREMREL. 02, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
 GN L1-CAM.
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=BRAIN, AND MUSCLE;
 RX MEDLINE=98147998; PubMed=9479034;
 RA Coutelle O., Nyakatura G., Taudien S., Elgar G., Brenner S.,
 RA Platzer M., Drescher B., Jouet M., Kenrick S., Rosenthal A.;
 RT "The neural cell adhesion molecule L1: genomic organisation and
 RT differential splicing is conserved between man and the pufferfish
 RT Fugu.";
 RL Gene 208:7-15(1998).
 CC - FUNCTION: CELL ADHESION MOLECULE WITH A AN IMPORTANT ROLE IN THE
 CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO AXONIN ON NEURONS.
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC - ALTERNATIVE PRODUCTS: TWO FORMS; BRAIN ISOFORM (SHOWN HERE) AND
 CC MUSCLE ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 6
 CC C2-LIKE DOMAINS. STRONG, TO DROSOPHILA NEUROGLIAN.
 CC - SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 DR EMBL; Z71926; CAA96469.1; -;
 DR EMBL; AF026198; AAC15580.1; -;
 DR HSSP; P20241; ICFB.
 DR InterPro; IPR001777; -;
 DR InterPro; IPR003006; -;
 DR Pfam; PF00041; fn3; 5.
 DR Pfam; PF00047; ig; 6.
 DR PRINTS; PR00014; FNTYPEIII.
 DR SMART; SM00060; FN3; 1.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
 KW Immunoglobulin domain; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 1277
 FT DOMAIN 30 1135
 FT TRANSMEM 1136 1156
 FT DOMAIN 1157 1277
 FT DOMAIN 52 146
 FT DOMAIN 147 252
 FT DOMAIN 253 349
 FT DOMAIN 350 441
 FT DOMAIN 442 533
 FT DOMAIN 534 631
 FT DOMAIN 632 728
 FT DOMAIN 729 831
 FT DOMAIN 832 932
 FT DOMAIN 933 1032
 FT DOMAIN 1033 1127
 FT DISULFID 173 224
 FT VARSPLIC 42 47
 FT YNSPLIC 1190 1193
 FT CARBOHYD 317 317
 FT CARBOHYD 503 503
 FT CARBOHYD 520 520
 FT CARBOHYD 531 531
 FT CARBOHYD 794 794
 FT CARBOHYD 839 839
 FT CARBOHYD 1035 1035
 FT CARBOHYD 1046 1046
 FT CARBOHYD 1068 1068
 FT CARBOHYD 1083 1083

FT CARBOHYD 1108 1108 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1277 AA; 141954 MW; 284BB49BA4A42C27 CRC64;
 Query Match 4.7%; Score 90.5; DB 13; Length 1277;
 Best Local Similarity 21.5%; Pred. No. 36;
 Matches 68; Conservative 38; Mismatches 116; Indels 95; Gaps 15;
 QY 31 EKFSASEIGPEATAFIGITFINEHIIENAPTRGHVSGSGNDPSLSYPLLIALLKDL 90
 Db 793 QNFAFEIKVQAVNK-KGL-----GPEDDPIIGSGEDVLEAPNLGVLLNS 840
 QY 91 TVVEVAAEHFYDGGKKWYIILTEAMKAVIRLALFRNSGYKMLQGGTPEEKDKNOS 150
 Db 841 TTRVT-----WSAV-----DKETVRGHLL---GKYILTWG---HHRNSRAQEP 879
 QY 151 QNRA-GNSGRNLGPHGLGNQNHHPNLEGRMSALSFQGN--ARTTSSSTGWSRRIQ 207
 Db 880 ENIVMVGOTGANEKKSTINLRPYCHYDL---AISAFKSGEGPLSEKTSFTWP----- 929
 QY 208 HQQAVIEPPMIKERRRTMSELLT-----EKGVNGALFAIG-----EVLYITR 249
 Db 930 --EGVPGPPMSQMTSPSESEITLHWTPPKPKNGLLGYSLQYKRMOSDDNPLQVVDIAS 987
 QY 250 PLIYVLFIRKYGVRSWIPWAI-----SLSDVTLGMLL-----ANS 285
 Db 988 FEITHLTLLKGLDRSHYQVLLMARTAAAGKGLSIEILGATTLLEGPPANISLSAEERSVNL 1047
 QY 286 KMWGEKSK-----QVHF 297
 Db 1048 SWEARKRHRTVGFQIH 1064
 RESULT 15
 Q9ZW09 PRELIMINARY; PRT; 627 AA.
 ID Q9ZW09
 AC Q9ZW09;
 DT 01-MAY-1999 (TREMREL. 10, Created)
 DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE PUTATIVE RECEPTOR-LIKE PROTEIN KINASE.
 GN F16P2.40
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
 RT "Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004561; AAC95213.1; -;
 DR HSSP; P05045; 1BQJ.
 DR InterPro; IPR000719; -;
 DR InterPro; IPR001220; -;
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00139; lectin_legB; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR SMART; SM00221; STYK; 1.
 KW ATP-binding; Kinase; Receptor; Transferase.
 SQ SEQUENCE 627 AA; 68788 MW; 91D9F8100BEIAB1 CRC64;
 Query Match 4.7%; Score 90; DB 10; Length 627;
 Best Local Similarity 20.8%; Pred. No. 14;
 Matches 66; Conservative 46; Mismatches 131; Indels 74; Gaps 11;
 QY 16 QSGFGFANGTLWLPKFSASEIGPEAVTAFIGITFINE-----HIEN 60
 Db 95 QGTGAPGHGLAFVIFSPDMDFSGAFP---SNYLGLENTSNGNSLNRLAIEFDTVQAVEL 151

QY 61 APTPRGHVSSGNDPSLSYPLLIALLKLETVVEVAA--EHFYGDKKKXKXIIITEAMKAV 118
Db 152 NDIDONHVGI-----DLNGVISIASAPAYFDDREAKNISLRLASGKP 194
QY 119 IRLALFRNSGYKMLLOGGETPNEEKDSNQSESONRAGNSGRNLGPHGLGNQNHHPWNLE 178
Db 195 VRVWIEYNATETMLNVTIAPLDRPKPSIPLLS-----RKMNLGIFSQEH----- 240
QY 179 GRAMSALSSFGQNARTTTSS--TPGWSRRIOHQA---VIEPPMIKERRRTMS-----EL 228
Db 241 -----VGFSASTGTVASSHVFLGWSFNIEGKESDFITKLPSPDPPTLSPPSPPV 293
QY 229 LTERGVNGALFAIGEVLVITRPLIYVLFIRKYGVR-----WIPWAISSLVDTL--GMGLL 282
Db 294 STEKKSNTMLIIIIAASATVALMILIFSGFWFLRRDKIFFIGGARKFSYQTISNATGGF 353
QY 283 ANSKWGEKSKQVHFG 299
Db 354 DNSKLLGERNSGSFYKG 370

Search completed: July 11, 2001, 11:37:04
Job time: 123 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 12.29 Seconds
(without alignments)
601.553 Million cell updates/sec

Title: US-09-545-072A-2

Perfect score: 1922

Sequence: 1 MEAYKQWNRNREYVQSGFS.....EKIVELLEGAQSRVYISGS 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89.5	4.7	1257	2	US-08-750-152A-2
2	89	4.6	3165	2	US-08-459-146-3
3	89	4.6	3165	2	US-08-459-065-3
4	88	4.6	1489	6	5183745-2
5	87.5	4.6	302	2	US-08-893-853-3
6	87.5	4.6	302	4	US-09-113-921-3
7	86.5	4.5	1012	2	US-08-475-891A-4
8	86.5	4.5	1025	2	US-08-567-375-4
9	86.5	4.5	1025	2	US-08-587-680A-4
10	85.5	4.4	2595	4	US-09-036-987A-2
11	81.5	4.2	1250	1	US-08-441-139-9
12	80.5	4.2	524	1	US-08-447-500-2
13	80.5	4.2	524	1	US-08-454-097-24
14	80.5	4.2	524	1	US-08-453-866-24
15	80.5	4.2	524	3	US-08-185-359-24
16	80.5	4.2	747	2	US-08-895-522-1
17	80.5	4.2	747	3	US-09-195-391-1
18	80	4.2	730	2	US-08-696-944-2
19	79.5	4.1	821	1	US-07-928-464-2
20	79.5	4.1	821	1	US-08-003-311B-2
21	79.5	4.1	821	1	US-08-261-432-2
22	79.5	4.1	821	5	PCT-US93-07347-2
23	79	4.1	1794	6	5183745-6
24	78.5	4.1	513	1	US-08-464-340A-2
25	78.5	4.1	513	5	PCT-US94-08449A-2
26	77	4.0	525	6	5183745-4
27	77	4.0	1119	4	US-09-396-651B-2

28	76.5	4.0	392	4	US-09-416-050A-2	Sequence 2, Appli
29	76.5	4.0	392	4	US-09-664-800-2	Sequence 2, Appli
30	76.5	4.0	392	4	US-09-665-309-2	Sequence 2, Appli
31	76.5	4.0	1989	4	US-08-836-325-12	Sequence 12, Appli
32	76.5	4.0	3079	5	PCT-US94-00198-4	Sequence 4, Appli
33	75	3.9	3025	6	5223423-3	Patent No. 5223423
34	74.5	3.9	307	2	US-08-540-804-6	Sequence 6, Appli
35	74.5	3.9	307	2	US-08-218-265-6	Sequence 6, Appli
36	74.5	3.9	307	3	US-08-521-872-6	Sequence 6, Appli
37	74.5	3.9	307	4	US-08-590-399-6	Sequence 6, Appli
38	74.5	3.9	1052	4	US-09-255-502-7	Sequence 7, Appli
39	74.5	3.9	1261	4	US-09-208-742-4	Sequence 4, Appli
40	74	3.9	494	1	US-08-447-500-2	Sequence 4, Appli
41	74	3.9	494	1	US-08-454-097-2	Sequence 2, Appli
42	74	3.9	494	1	US-08-447-408-2	Sequence 2, Appli
43	74	3.9	494	1	US-08-453-866-2	Sequence 2, Appli
44	74	3.9	494	3	US-08-185-359-2	Sequence 2, Appli
45	73.5	3.8	230	4	US-09-091-219-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-750-152A-2
; Sequence 2, Application US/08750152A
; Patent No. 5977331
; GENERAL INFORMATION:
; APPLICANT: ASAKURA, YOKO
; APPLICANT: KIMURA, EIICHIRO
; APPLICANT: ABE, CHIZU
; APPLICANT: KAWAHARA, YOSHIO
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: ALPHA-KETOGLUTARATE DEHYDROGENASE GENE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22152
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,152A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBOLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-750-152A-2

Query Match 4.7%; Score 89.5; DB 2; Length 1257;
Best Local Similarity 20.2%; Pred. No. 1.3;
Matches 62; Conservative 37; Mismatches 123; Indels 85; Gaps 13;

```
QY 65 RGVSSGNDPSLSYPLLIATLKDLTETVEVAABHFYDGGKKNWYIILTEAMKAVIRLALF 124
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 781 RGH--NEADPSMTQPKMYELITRETVAQYTEDLLGRGD-----LSNEDAEAVR--DF 832
QY 125 RNSGKMLLOGGETPNEEKSDNQESQNRAGNSGRNLGPHGLGNO-----169
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 833 HD-----QMESVNEVKGKQAEQAQTGIGTSQKLPGLHLETNISREELLEGQAFAN 885
QY 170 -----NHNH-----PWNLEGRAMSALSFGQNRATTTSTSPGWSR 204
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 886 TPEGFNHPRVAPVAKKRVSSVTBGGIDWAG-ELLAFGLANSRLVRLAGED-----SR 940
QY 205 R--IQHQAVTEPMIXERRTHSELTEKGVNGALFAIGEVLVITRPLIYVIFIRKYG 262
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 941 RGTFTQRAVAIDPATAEFNPLHQAQSKNGKF-----LVYNSALTEYAGMGFEYGY 995
QY 263 R-----SWIPWALSLSVDTLGMGLL-----ANSKMWGEKSKOV-----HPSGPEKDE 304
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 996 SVGNEDSVVAWEAQFGDFANGAQTIIDEYVSSGEAKWGQTSKLILLPHGYEGQGDHSS 1055
QY 305 LRRKLI 311
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 1056 ARIERFL 1062

RESULT 2
US-08-459-146-3
; Sequence 3, Application US/08459146
; Patent No. 5866405
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,146
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3165 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryptonectria
; ORGANISM: parasitica)
; STRAIN: EP713
; US-08-459-146-3
```

Query Match 4.6%; Score 89; DB 2; Length 3165;

Best Local Similarity 20.3%; Pred. No. 6.2;

Matches 82; Conservative 49; Mismatches 149; Indels 124; Gaps 18;

```
QY 4 YKOW--VMNRREYVOSFGSFANGTLWLLPEKFSASEIGPEAVTAF-----LGIFTT 52
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 1757 FEQWPEVFADRDIM-----LPKGVELYIKESYA---GTPFISSEYKSRKALKQAGVMDV 1808
QY 53 INEHIENAPTRGHVSSGNDPSLSY-----PLLIALLKOLETVVEVAABHF 100
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 1809 IKNALECT-----STGKYPTQFYHAFKSAQVPGOPLLAPRMKDLRTVVSDDL 1860
QY 101 YGDKKNWYIILTEAMKAVIRLALFRNSGVKMLLOGGETPNEEKSDNQESQN-----RAGN 156
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 1861 MYDQ-----IFQTEANKRITWETYGAGSGMPLSQSMARTWDELHDKRREGGQFIADATA 1916
QY 157 SGRNLGP---HG-----LGNQNHHPNWNLEGRAMSALSSEFG-----QNARTTTSTPGWS 203
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 1917 YDSNCKPALFHGAGKLVGLGFGNHPSG---KGRQFAQVVOCKFEAMQNAWYMGITEPSYT 1973
QY 204 RRIHQQAQVIEPPMIKERR-----RTWSELTEKGVNGALFAIGEVLVITRPLIYVLFTRK 259
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 1974 ALTFHVPDVAVRHELESKYPAHFATFSELLAHNNVN-----VTE 2012
QY 260 YGVRSW-----IPWALSLSVDTLGMGLLANSKMWGEKSKOVHFSGPEKDELRR 307
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 2013 WKRLSWEERKACARDMQAVPGKVFITNDPALR--LQSSWGSQSFTE-----PKRDE--- 2062
QY 308 RKLIALYLMRDPFTKYTRQKLESSQKKLELIPLIGLFTLEKIV 351
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 2063 -----PRKYOTVYFDSKAAAREDIKRIVFANREVI 2092
```

```
RESULT 3
US-08-459-065-3
; Sequence 3, Application US/08459065
; Patent No. 5882642
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,065
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 3165 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryphonectria)
; STRAIN: EP713
; US-08-459-065-3

Query Match 4.6%; Score 89; DB 2; Length 3165;
Best Local Similarity 20.3%; Pred. No. 6.2;
Matches 82; Conservative 49; Mismatches 149; Indels 124; Gaps 18;

QY 4 YKOW--VVRNREYVQSGFANGLTWLLPEKFSASEIGPEAVTAF-----LGIFTT 52
Db 1757 FEQPEVFADRDIM-----LPKGVELYIKESYA--GTFPFISSFYKSRKALKQAGYMDV 1808
QY 53 INEHIENAPTGRGHVGGNDPSSLSY-----PLLIALLKDLTVEVAAEHF 100
Db 1809 IRKNALECI-----STCKYPTQFYHAFKSAQVPGQPLLAPRMKDLRTVVSSELSAY 1860
QY 101 YGDKKNYIILTEAMKAVIRLALFRNSGYKMLQGGTPEEKDSNSESQ-----RAGN 156
Db 1861 MVDO-----IFQIEANKRITWETYGAGSGMPLSQSMARIMDELHDLKRGQGFIIADATA 1916
QY 157 SGRNLGP---HG-----LCNONHNNPNLEGRAMSALSSFG---QNATTTSTSGWS 203
Db 1917 YDSNCKPALFHGAGKLVEFGQNHPSG---KGROFAOVVOCKFEAMQNAVVMGITEPSYT 1973
QY 204 RRIHQOQAVIEPPMIKERR-----RTMSELLTEKGVNGALFAIGEVLVITRPLIYVLFIRK 259
Db 1974 ALTPHVDVAVRHELESKYPAHFATPSELLAHNNVN-----VTE 2012
QY 260 YGVRSW-----IPWALSVDTLGMGLLANSKWKGEKSKOVHFGSGPEKDLRR 307
Db 2013 WKRLSWEERKACARDMQAVFGKVFELTNDPALR---LOGSSWQSGFTTE-----PKRDE--- 2062
QY 308 RKLIALYLMRDPFTKYTRQKLESSQKLELIPLIGLFTKIV 351
Db 2063 -----FRKYQYFYDSKAMREDIKRIVFANREVI 2092

RESULT 4
; Patent No. 5183745
; APPLICANT: DANCHIN, ANTOINE; CLASER, PHILLIPPE; KRIN, EVELYN;
; BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
; BIOLOGICAL USES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,541
; FILING DATE: 25-OCT-1989
; SEQ ID NO: 2
; LENGTH: 1489
5183745-2

Query Match 4.6%; Score 88; DB 6; Length 1489;
Best Local Similarity 21.4%; Pred. No. 2.4;
Matches 69; Conservative 40; Mismatches 105; Indels 108; Gaps 16;

QY 25 LTWLLPEKFSASEIGPEAVTAF-----LGIFTINEHIENAPTGRGHV----- 68
Db 241 LLWKIARAGARSAYGTEARROFYDGMNIGVITDF--ELEVNRNALNRRRAHVAAGQDVVQH 299
QY 69 GSSGNDPSSLSYPLLIIALLKDLTVEVAAEHFYGDKKNYIILTEAMKAVIRLALFRNSG 128
Db 300 GTEONPN---FP-----EAEKIFVVSATGESG 324

; LENGTH: 3165 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryphonectria)
; STRAIN: EP713
; US-08-459-065-3

Query Match 4.6%; Score 89; DB 2; Length 3165;
Best Local Similarity 20.3%; Pred. No. 6.2;
Matches 82; Conservative 49; Mismatches 149; Indels 124; Gaps 18;

QY 4 YKOW--VVRNREYVQSGFANGLTWLLPEKFSASEIGPEAVTAF-----LGIFTT 52
Db 1757 FEQPEVFADRDIM-----LPKGVELYIKESYA--GTFPFISSFYKSRKALKQAGYMDV 1808
QY 53 INEHIENAPTGRGHVGGNDPSSLSY-----PLLIALLKDLTVEVAAEHF 100
Db 1809 IRKNALECI-----STCKYPTQFYHAFKSAQVPGQPLLAPRMKDLRTVVSSELSAY 1860
QY 101 YGDKKNYIILTEAMKAVIRLALFRNSGYKMLQGGTPEEKDSNSESQ-----RAGN 156
Db 1861 MVDO-----IFQIEANKRITWETYGAGSGMPLSQSMARIMDELHDLKRGQGFIIADATA 1916
QY 157 SGRNLGP---HG-----LCNONHNNPNLEGRAMSALSSFG---QNATTTSTSGWS 203
Db 1917 YDSNCKPALFHGAGKLVEFGQNHPSG---KGROFAOVVOCKFEAMQNAVVMGITEPSYT 1973
QY 204 RRIHQOQAVIEPPMIKERR-----RTMSELLTEKGVNGALFAIGEVLVITRPLIYVLFIRK 259
Db 1974 ALTPHVDVAVRHELESKYPAHFATPSELLAHNNVN-----VTE 2012
QY 260 YGVRSW-----IPWALSVDTLGMGLLANSKWKGEKSKOVHFGSGPEKDLRR 307
Db 2013 WKRLSWEERKACARDMQAVFGKVFELTNDPALR---LOGSSWQSGFTTE-----PKRDE--- 2062
QY 308 RKLIALYLMRDPFTKYTRQKLESSQKLELIPLIGLFTKIV 351
Db 2063 -----FRKYQYFYDSKAMREDIKRIVFANREVI 2092

RESULT 4
; Patent No. 5183745
; APPLICANT: DANCHIN, ANTOINE; CLASER, PHILLIPPE; KRIN, EVELYN;
; BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
; BIOLOGICAL USES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,541
; FILING DATE: 25-OCT-1989
; SEQ ID NO: 2
; LENGTH: 1489
5183745-2

Query Match 4.6%; Score 87.5; DB 2; Length 302;
Best Local Similarity 24.1%; Pred. No. 0.22;
Matches 45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;

QY 64 PRGHVGGSGNDPSSYPLLIIALLKDLTVEVAAEHFYGDKKNYIILTEAMKAVIRLAL 123
Db 91 PWEHPGSSGVDPRL-----EPWNHL-----GSSGVDRHLEP 121
QY 124 FRNSGYKMLQGGTPEE-----KDSNSESQNRAGNSG---RNLPGLHGN-QN 170
Db 122 WKHPGSGDLGRRTTPQDSSGRRRPPQDSSGRRRPPQDSSGRRRPPQDSSGRRRPPQDSSGRR 181
QY 171 HHNPWNLEG-----RAMSALSFGONARTTSTPCWSRRRICHQOQAVIEPP 216
Db 182 RRPQNSSGRRRRSPQDSSGRRRRSPQDSSGRRRRSPQDSSGRRRRSPQDSSGRRRRRAHNS- 236
```

Query Match	4.5%;	Score 86.5;	DB 2;	Length 1012;
Best Local Similarity	23.9%;	Pred. No. 1.9;		
Matches	93;	Conservative	50;	Mismatches 145; Indels 101; Gaps 24;
QY	8	VWRNREYVQSGFSPANGLTWLLPEKFSASEIGPEAVATFAGTFTINEHILE--NAPTPR	65	
DB				
QY	244	IW- NLSIRAFSVRENKLGMP-----TWAFKTL--HLLEVIDMGNTNR	284	
DB				
QY	66	GHVGSSNDPFSYPLLIATLKDLTVEVAAAEHFYGDKKWNYILTEA---MKAVIRLA	122	
DB				
QY	285	FHG-----KIPASVANASHL-TVIQIVGNLFGS-----LIISGFGRLRNLTLEY	327	
DB				
QY	123	LFRNSGYKMLLOGGETPNEEKDSNQSSQNRAGNSGNRLGPHGLGICGNHHNPNWNLGRAM	182	
DB				
QY	328	LWRN-----LFQ---TREQDDWGFISDLTNCISKIQTNLGNENNLGGVLPNFSNLS	375	
DB				
QY	183	SALSSFGONARTTTSTSP---GWSRRHQHOQAVIEP-----PMIKERRRMTSELLT-EKG	233	
DB				
QY	376	TSLSFALENKLTGSLPKDIGNLIGLQHLYLCNNNFRGSLPSSLGRLKNGILGILLAYENN	435	
DB				
QY	234	VNGAL-FAIGEVILYTRPLIYVIFIRKYGVRSWPWAISSLVDTLGMGLKNSKWWGEKS	292	
DB				

Db 436 LSGSIPLAIGN---LTELNILLGTNKFSGWIPYTLNLTNLLSLGLSTN----- 482

QY 293 KOVHFSGPEKDELRRKLIWALYLMRDPFTKYTRQKLESS--QKKLELPLIGF----- 345

Db 483 ---NLSGPIPSSELFN---IOTLSIM-----INVSNNLEGSIPQEIGHLKNLVEFHAESN 531

QY 346 -LTEKIVELLEGAGQ-SRYTY-----ISGS 367

Db 532 RLSGKIPNTLGDQCLLYLYLQNNLLSGS 560

RESULT 8

US-08-567-375-4

; Sequence 4, Application US/08567375

; Patent No. 5952485

; GENERAL INFORMATION:

; APPLICANT: Ronald, Pamela C.

; APPLICANT: Wang, Guo-Liang

; APPLICANT: Song, Wen-Yuang

; APPLICANT: Szabo, Veronique

; TITLE OF INVENTION: Procedures and Materials for Conferring

; TITLE OF INVENTION: Disease Resistance in Plants

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/567,375

; FILING DATE: 04-DEC-1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/004,645

; FILING DATE: 29-SEP-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/475,891

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/373,375

; FILING DATE: 17-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Bastian, Kevin L.

; REGISTRATION NUMBER: 34,774

; REFERENCE/DOCKET NUMBER: 023070-058930

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1025 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-567-375-4

Query Match 4.5%; Score 86.5; DB 2; Length 1025;

Best Local Similarity 23.9%; Pred. No. 2;

Matches 93; Conservative 50; Mismatches 145; Indels 101; Gaps 24;

QY 8 VWRNREYVQSGFANGITWLLPEKFSASEIGPEAVTAFLGIFTTNEHIE--NAPTPR 65

Db 244 IW-NLSSLRASFVRENKLGMP-----INAEKTL--HLEVIDMGTR 284

QY 66 GHVGSSGNDPSLSYPLLIALLKDLTVVVAEHHFYGDKKNYIILTEA---MKAVIRLA 122

Db 285 FHG-----KIPASVANASHL-TVIQYGNLFSG-----IITSGFRLRLNLTLY 327

QY 123 LFRNSYKMLLQGGTPEPNEKDSQSEONRAGNSGRNLGPHCLGNQHHNPWNLEGRAM 182

Db 328 LWRN-----LFQ---TFREDDWGFISDLTNCCKLOPLNGLNENLGGVLPNSFSNLS- 375

QY 183 SALSSEFGQARTTTSTP---GMSRRHQQAVIDP-----PMIKERRRTMSELIT-EKG 233

Db 376 TSLSLALELNLKITGSIKPRDIGNLIGLQHLLYLCNNFRGSLPSSGLRKNLGLLAYENN 435

QY 234 VNGAL-FAIGEVLYITRPLIYVIRKYGVRSWIPRAISLVSVDTLGMLLANSKWWGEKS 292

Db 436 LSGSIPLAIGN---LTELNILLGTNKFSGWIPYTLNLTNLLSLGLSTN----- 482

QY 293 KOVHFSGPEKDELRRKLIWALYLMRDPFTKYTRQKLESS--QKKLELPLIGF----- 345

Db 483 ---NLSGPIPSSELFN---IOTLSIM-----INVSNNLEGSIPQEIGHLKNLVEFHAESN 531

QY 346 -LTEKIVELLEGAGQ-SRYTY-----ISGS 367

Db 532 RLSGKIPNTLGDQCLLYLYLQNNLLSGS 560

RESULT 9

US-08-587-680A-4

; Sequence 4, Application US/08587680A

; Patent No. 5977434

; GENERAL INFORMATION:

; APPLICANT: Ronald, Pamela C.

; APPLICANT: Wang, Guo-Liang

; APPLICANT: Song, Wen-Yuang

; APPLICANT: Szabo, Veronique

; TITLE OF INVENTION: Procedures and Materials for Conferring

; TITLE OF INVENTION: Disease Resistance in Plants

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/587,680A

; FILING DATE: 17-JAN-1996

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/373,375

; FILING DATE: 17-JAN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/475,891

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/004,645

; FILING DATE: 29-SEP-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/567,375

; FILING DATE: 04-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Bastian, Kevin L.

; REGISTRATION NUMBER: 34,774

; REFERENCE/DOCKET NUMBER: 023070-058940US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

;> LENGTH: 1025 amino acids
;> TYPE: amino acid
;> TOPOLOGY: linear
;> MOLECULE TYPE: protein
US-08-587-680A-4

Query Match 4.5%; Score 86.5; DB 2; Length 1025;
Best Local Similarity 23.9%; Pred. No. 2;
Matches 93; Conservative 50; Mismatches 145; Indels 101; Gaps 24;

QY 8 WVRNREYVQSGSFANGLTWLLPEKFSASEIGPEAVTAFLGIFTTINEHIETNAPTTPR 65
DB 244 IW-NLSSRAFSVRKNKGMP-----TNAFKTL--HLEVIDMGNTNR 284
QY 66 GHVGSSGNDPSLSYPLLIAILKDLTVEVAAAEHFYGDKNWYIILTEA---MKAVIRLA 122
DB 285 FHG-----KIPASVANASHL-TVIQIYGNLPSG-----IISGGRRLNLTLY 327
QY 123 LFRNSGYKMLQGGTPENEKDSNQESQNRAGNSGRNLGPHGLGNGNHHNPNWNLGRAM 182
DB 328 LWRN-----LFQ-----TREDDMGFISDITNGSKLQTLNLGNNLGGVLPNSFSNLS--- 375
QY 183 SALSSFGQARTTTSSYP-----GWSRRHQHQAVIEP-----PMIKERRTMSSELLT-EKG 233
DB 376 TSLFLALELNKITSIPKIDIGNLIGLOHLYLCNNNFRGSLPSSLGRLKNLIGILLAYENN 435
QY 234 VNGAL-FAIGEVLYITRPLIYVIRKYGRSWIPWALSISVDPLGMLLANSKWKWGEKS 292
DB 436 LSGSIPLAIGN---LTELINILLGTNKS--GWIPYTLNLTLLSLGLSTN----- 482
QY 293 KQVHFSGPEKDLRRRLINWALYMRDPFTKTKROKLESS--QKKLELPLIGF----- 345
DB 483 ---NLSGPISSELEN---IQTLSTM-----INVSNNLEGSIPQEIHLKNLVEFHAESN 531
QY 346 -LTEKIVELLEGAQ-SRITY-----ISGS 367
DB 532 RLSGKIPTLGDCLLRLYLYLQNNLLSGS 560

RESULT 10
US-09-036-987A-2
;> Sequence 2, Application US/09036987A
;> Patent No. 6143526
;> GENERAL INFORMATION:
;> APPLICANT: Baltz, Richard H.
;> APPLICANT: Broughton, Mary C.
;> APPLICANT: Crawford, Kathryn P.
;> APPLICANT: Madduri, Krishnamurthy
;> APPLICANT: Merlo, Donald J.
;> APPLICANT: Treadway, Patti J.
;> APPLICANT: Turner, Jan R.
;> APPLICANT: Waldron, Clive
;> TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;> TITLE OF INVENTION: Production
;> NUMBER OF SEQUENCES: 39
;> CORRESPONDENCE ADDRESS:
;> ADDRESS: Dow Agrosciences LLC Patent Department
;> STREET: 9330 Zionsville Road
;> CITY: Indianapolis
;> STATE: Indiana
;> COUNTRY: USA
;> ZIP: 46268
;> COMPUTER READABLE FORM:
;> MEDIUM TYPE: Floppy disk
;> COMPUTER: IBM PC compatible
;> OPERATING SYSTEM: PC-DOS/MS-DOS
;> SOFTWARE: PatentIn Release #1.0, Version #1.30
;> CURRENT APPLICATION DATA:
;> APPLICATION NUMBER: US/09/036,987A
;> FILING DATE: 09-MAR-1998
;> CLASSIFICATION: 435
;> ATTORNEY/AGENT INFORMATION:

;> NAME: Stuart, Donald R
;> REGISTRATION NUMBER: 28,479
;> REFERENCE/DOCKET NUMBER: 50,608
;> TELECOMMUNICATION INFORMATION:
;> TELEPHONE: (317)337-4816
;> TELEFAX: (317)337-4847
;> INFORMATION FOR SEQ ID NO: 2:
;> SEQUENCE CHARACTERISTICS:
;> LENGTH: 2595 amino acids
;> TYPE: amino acid
;> TOPOLOGY: linear
;> MOLECULE TYPE: protein
US-09-036-987A-2

Query Match 4.4%; Score 85.5; DB 4; Length 2595;
Best Local Similarity 19.3%; Pred. No. 11;
Matches 78; Conservative 51; Mismatches 138; Indels 137; Gaps 17;
QY 7 WVRNREYVQSGSFANGLTWLLP-EKFSASEIGPEAVTAFLGIFTTINEHIETNAPTTPR 65
DB 1766 WRQVREPVR---PADGVOALVEHDVATVVELGPDGALSAL-----IQECVAAS 1811
QY 66 GHVGSSGNDPSLSYPLLIAILKDLTVEVAAAEHFY---GDKKW-NYIILTEAMKAVIRL 121
DB 1812 DHAGRLSAVPMRRN-----QDEAKVMTALAHVVRGAVDWSRFFAGTRAKQIELPT 1865
QY 122 ALFRNSGYKMLQGGTPENEKDSNQESQNRAGNSGRNLGPHGLGNGNHHNPNWNL---- 177
DB 1866 YAFQQRQRYWL-----NALRESS-----AGDMGRVREAKFWGAVEHEDVESLARVL 1910
QY 178 ---EGRAMSALSSSF-----QGNARTTTS----- 197
DB 1911 GIVDDGAAVDSLSALPVLAGHQRTTETESINDPCRYLGRVQVAGLPPMGTVFTWLVF 1970
QY 198 STPGHSR-----RIHQQAVI-----EPPMIKERRRT-----MSELLT 230
DB 1971 APHGSSEPEVVDCVTALRARGASVYLVEADPDPTSFQDRVTRTLCGLPDLVGVLMLCL 2030
QY 231 EKGVNGALFAIGEVLYITRPLIYVIRKYGRSWIPWALSISV-----DTLGMCL- 281
DB 2031 EESVLPGFSAVSRGFALTVELVRLRAAGATARLMLLTCCGGVSGVDVPRPAQALAWGLG 2090
QY 282 -----LANSKWG-----EKSQVHFSGPEKDELRRR 308
DB 2091 RVVGLHPDMWGLIDIPVLEDEDAQERLSIVLAGLDEDEVAIR 2134

RESULT 11
US-08-441-139-9
;> Sequence 9, Application US/08441139
;> Patent No. 5773245
;> GENERAL INFORMATION:
;> APPLICANT: Wittrup, Dr. Karl D.
;> APPLICANT: Robinson, Anne S.
;> TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
;> TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
;> NUMBER OF SEQUENCES: 20
;> CORRESPONDENCE ADDRESS:
;> ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
;> STREET: 400 Garden City Plaza
;> CITY: Garden City
;> STATE: NY
;> COUNTRY: USA
;> ZIP: 11530
;> COMPUTER READABLE FORM:
;> MEDIUM TYPE: Floppy disk
;> COMPUTER: IBM PC compatible
;> OPERATING SYSTEM: PC-DOS/MS-DOS
;> SOFTWARE: PatentIn Release #1.0, Version #1.25
;> CURRENT APPLICATION DATA:
;> APPLICATION NUMBER: US/08/441,139
;> FILING DATE: 15-MAY-1995


```

CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/089,997
FILING DATE: 06-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8646
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1250 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-441-139-9

Query Match          4.2%; Score 81.5; DB 1; Length 1250;
Best Local Similarity 21.9%; Pred. No. 9.7;
Matches 68; Conservative 34; Mismatches 115; Indels 93; Gaps 14;

QY      31 EKFSASEIGP-----EATFAELGLFTTINEHIENAPTGRGHVG-----SSGNDPSLSVPL 81
DB      180 KRISGGAGVTVKIDSVDADSGSNSTEQSDFKFPNNAHOGHRRATNSLSPPSFKPP- 238
QY      82 LIALKDLTVEVAABHFYGDKKWNIILTEAMKAVIRLALFRNSGYKMLLQGGETENE 141
DB      239 -----PNSHGDNDEFI---ATSTHRRSKTRNEY----SPGINSNW 274
QY      142 EKDSNQSESQ-----NRAGSNGRNLPGLGNQHHPNPLEGRAMSALSFGQN 191
DB      275 RNQSQQPOQLSPRHRGSNSROYNSFNTLEPAIFQOQHKH--RASNSSVHSFSSQGN 332
QY      192 ARTTSSTPGWSRIQH----QQAIVLEPPMIKERRTMSSELLTEK-----GVNGALF 239
DB      333 -----NGGRKSLEFAPYLPOANI-PETIQEGRVLVAGILRVKNKRSDAWSTDGALD 383
QY      240 AIGEVLXITREPLYLVLFIRKYGVRSWIPWAISVDTLGMGLIANSKWKGEKSKOVHFSG 299
DB      384 AD-----IYICGSKDRN-----RALEDGLVAVELIVDDVWEKK----- 418
QY      300 PEKDLELRKK 309
DB      419 -EKEEKRRK 427

RESULT 12
US-08-447-500-24
; Sequence 24, Application US/08447500
; Patent No. 5627064
; GENERAL INFORMATION:
; APPLICANT: Hoeckstra, Merl F.
; TITLE OF INVENTION: PROTEIN KINASES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,500

```


Db 351 -----LHGYGNPRVNGNTARNVNTNSKTRNTTPVATPKQQAQNSYNKD----- 396
QY 188 FGQNAFTTTSTPGWSRRRIHQHQAQVIEP 215
Db 397 ---NSKSRISSNPQSFTKQQHVLKKIEP 421

Search completed: July 11, 2001, 11:36:34
Job time: 93 sec

